A REGRESSION DESIGN APPROACH TO OPTIMAL AND ROBUST SPACING SELECTION

by

Randall L. Eubank

Technical Report No. 144
Department of Statistics ONR Contract

July 1981

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

A Regression Design Approach to Optimal and Robust Spacing Selection

By R. L. Eubank

Short title: Spacing Selection

Summary. The problem of location and/or scale parameter estimation using the asymptotically best linear unbiased estimator based on sample quantiles is considered. The problem of optimal spacing selection for these estimators is shown to be equivalent to the problem of regression design for time series with Brownian motion or Brownian bridge covariance structures and a particular variable knot spline approximation problem. This equivalence is employed, in conjunction with a regression framework, to investigate the asymptotic properties of certain spacing selection schemes. In particular, an asymptotic alternative is developed to a robust estimation procedure suggested by Chan and Rhodin (1980).

1. Introduction.

In a location and scale parameter model it is assumed that a random sample $X_1,\ldots,X_N$ is obtained from a distribution of the form

$$F(x) = F_0(\frac{x-\mu}{\sigma}) ,$$

where F_0 is a known distributional form and μ and σ are, respectively, a location and scale parameter. Usually μ and/or σ are unknown and must be estimated from the data. In this paper, the properties of the asymptotically best linear unbiased estimators (ABLUE's) of μ and σ based on $n < N$ sample quantiles will be investigated.

The ABLUE is an easily computed estimator which derives from the asymptotic distribution of the sample quantiles and was first suggested by Mosteller (1946). Further properties and computational formulas were latter derived by Ogawa (1951). In particular, Ogawa obtained explicit expressions for the asymptotic relative efficiency (ARE) of the ABLUE with respect to the Cramér-Rao lower variance bound for unbiased parameter estimation. The ABLUE has also received attention in the context of robust estimation due to the work of Chan and Rhodin (1980).

Define the sample quantile function, $\tilde{Q}$, by

$$\tilde{Q}(u) = X_{(j)}, \quad \frac{j-1}{N} < u \leq \frac{j}{N}, \quad j = 1,\ldots,N, \tag{1.1}$$

where $X_{(j)}$ denotes the jth sample order statistic. Then, for any set of real numbers $0 < u_1 < \ldots < u_n < 1$, the ABLUE's of μ and σ will have the form $\Sigma_{i=1}^{n} b(u_i)\tilde{Q}(u_i)$ (c.f. Ogawa (1951) or Eubank (1979) for explicit expressions for the $b(u_i)$ in the various estimation situations). Since the ABLUE uses a subsample of $n < N$ of the sample quantiles or order statistics, the quantiles which are utilized in the estimator, or equivalently their spacing, $u_1,\ldots,u_n$, must be chosen appropriately. The problem of optimally selecting the u_i, $i=1,\ldots,n$, has classically been termed the optimal spacing problem and has been addressed by Bloch (1966), Balmer, Boulton and Sack (1974), Chan (1970), Chernoff (1971), Eisenberger and Posner (1965), Gupta and Gnanadesikan (1966), Hassanein (1968, 1969a, 1969b, 1971, 1972, 1977), Kulldorf (1963), Kulldorf and Vännman (1973), Rhodin (1976), Sarhan and Greenberg (1958, 1962) and Särndal (1962, 1964).

The usual approach to the optimal spacing problem has been to attempt to find a spacing which corresponds to the maximum value for one of the

ARE expressions given by Ogawa. Whereas this is usually a straightforward, albeit tedious, numerical problem in the event of a single unknown parameter, the case when both parameters are unknown usually proves to be both analytically as well as numerically intractable. This latter fact has led to the use of "suboptimal" spacing selection schemes such as the selection of a spacing which maximizes the sum of the two ARE's of the estimators or a spacing which minimizes the sum of the estimators' variances. This type of approach has been employed by Eisenberger and Posner (1965) and Hassanein (1969a, 1969b, 1977).

In this paper the asymptotic (as $n \to \infty$) properties of optimal as well as suboptimal spacing selection schemes are derived and then utilized to obtain an analytic approach to robust estimation problems such as that considered by Chan and Rhodin (1980). These results are obtained using the regression analysis framework developed in Parzen (1979). When viewed in a regression setting, so called suboptimal spacings which, for instance, minimize the sum of the variances are seen to be well motivated from the perspective of regression design as well as for other reasons such as computational simplicity.

The asymptotic implication of several spacing selection criteria are considered in Section 3. The regression framework from which these results derive is presented in Section 2 where it is shown that the problem of optimal spacing selection can be viewed as both an optimal regression design problem and a variable knot spline approximation problem for splines of order 1. This fact permits the derivation of the asymptotic results in Section 3. Finally, in Section 4, the results of Section 3 are utilized to develop a general approach to a problem of robust ABLUE construction.

2. Regression design, spacing selection, and variable knot spline approximation. The objective of this section is to show the equivalence of three problems: the optimal spacing problem, the problem of regression design in the presence of errors with Brownian bridge (or Brownian motion) covariance structure, and the problem of finding the best $L^2[0,1]$ approximation of a particular function by piecewise constants with free breakpoints. First, however, a few preliminaries are required.

Now, and in subsequent discussions, it will be assumed that F_0 is absolutely continuous with associated density $f_0 := F_0'$ (where: = means "is defined as"). The quantile function corresponding to F_0 is $Q_0(u) := F_0^{-1}(u) := \inf_x \{x | F_0(x) \geq u\}$ and the density-quantile function is defined as $d_0(u) := f_0(Q_0(u))$, $0 \leq u \leq 1$.

Parzen (1979) has shown that for large N the problem of location and/or scale parameter estimation can be considered as a continuous parameter time series regression problem through use of the model

$$d_0(u)\tilde{Q}(u) = \mu d_0(u) + \sigma d_0(u)Q_0(u) + \sigma_B B(u) \quad , \; u \in [0,1] \; , \tag{2.1}$$

where $\sigma_B = \sigma/\sqrt{N}$ and $B(\cdot)$ is a Brownian bridge process, i.e., $B(\cdot)$ is a zero mean normal process with covariance kernel

$$R(u_1,u_2) = u_1 - u_1u_2, \quad 0 \leq u_1 \leq u_2 \leq 1 \; . \tag{2.2}$$

Consequently, under the regularity condition that both d_0 and the product of d_0 and Q_0, $d_0 \cdot Q_0$, are in the reproducing kernel Hilbert space (RKHS) generated by R, the techniques developed by Parzen (1961a, 1961b) may be utilized to construct linear estimators of μ and σ which are based on the entire set of N sample quantiles. Denote these estimators by $\hat{\mu}$ and $\hat{\sigma}$. Their corresponding variance-covariance matrix is $\sigma_B^2 A^{-1}$, where A is the usual intrinsic accuracy matrix associated with the location and scale parameter model (2.1).

The problem of optimal regression design selection for model (2.1) is also, clearly, a problem of optimal quantile selection. It is, in fact, the optimal spacing problem. To see this define the set of all possible n-point designs for model (2.1) by

$$D_n := \{(u_1,\dots,u_n) | 0 < u_1 < u_2 < \dots < u_n < 1\} .$$

Given a particular design, $U = \{u_1,\dots,u_n\} \varepsilon D_n$, the observation set $\{d_0(u_1)\tilde{Q}(u_1),\dots,d_0(u_n)\tilde{Q}(u_n)\}$ may be utilized, as a result of model (2.1), to construct estimators, $\mu(U)$ and $\sigma(U)$, for μ and σ through the use of generalized least squares. Denote the variance-covariance matrix of these estimators by $\sigma_B^2 A(U)^{-1}$. It has been noted by Eubank (1981) that $\mu(U)$ and $\sigma(U)$ coincide with the ABLUE's for μ and σ based on the spacing U and that $\sigma_B^2 A(U)^{-1}$ coincides with their asymptotic variance-covariance matrix. Since the ARE expression for simultaneous parameter estimation given by Ogawa (1951) is

$$ARE(\mu(U),\sigma(U)) = |A(U)|/|A| ,$$

where $|\cdot|$ denotes the determinant function, it is now apparent that the optimal spacing problem is identical with the problem of D-optimal design selection for model (2.1). The criterion of minimizing the sum of the estimators' variances is now recognized as A-optimal design selection since

$$\sigma_B^2 trA(U)^{-1} = Var(\mu(U)) + Var(\sigma(U)),$$

where tr denotes the matrix trace. Similarly, maximizing the sum of the ARE's is equivalent to maximizing $tr[A(U)M]$, where $M^{-1} := diag(a_{11},a_{12})$ and a_{ii}, $i=1,2$, denote the diagonal elements of A.

In the case of, for instance, D-optimal designs for model (2.1) it suffices to maximize $|A(U)|$ over $U \varepsilon D_n$. It should be noted that such a design is also D-optimal for the regression model

$$Y(t) = \beta_1 d_0(t) + \beta_2 d_0(t) Q_0(t) + X(t),\ t\varepsilon[0,1], \tag{2.3}$$

where $X(\cdot)$ is a Brownian bridge process. Similar remarks hold for other optimality criteria and for the case of only one unknown parameter. Thus, if $\mu(\sigma)$ is known an optimal spacing for estimating $\sigma(\mu)$ is also an optimal design for the estimation of $\beta_2(\beta_1)$ when $\beta_1(\beta_2)$ is known.

The problem of optimal design (and hence optimal spacing) selection may be analyzed using the RKHS approach developed by Sacks and Ylvisaker (1966, 1968). Therefore, let $H(R)$ denote the RKHS generated by R in (2.2) with associated norm denoted by $||\cdot||_R$. It can be shown (c.f. Parzen (1979)) that

$$H(R) = \{f|f(0)=f(1)=0,\ f'\varepsilon L^2[0,1]\}\ .$$

The inner product of $f,g\ \varepsilon\ H(R)$ is

$$\langle f,g\rangle_R = \int_0^1 f'(x)g'(x)dx = \langle f',g'\rangle_{L^2}\ , \tag{2.4}$$

where $\langle\cdot,\cdot\rangle_{L^2}$ denotes the usual $L^2[0,1]$ inner product. It now follows from the work of Sacks and Ylvisaker (1968) that the matrices A and $A(U)$ associated with the estimators $(\hat{\mu},\hat{\sigma})^t$ and $(\mu(U),\ \sigma(U))^t$ respectively, are given by

$$A = \begin{bmatrix} ||d_0||_R^2 & \langle d_0, d_0\cdot Q_0\rangle_R \\ \langle d_0, d_0\cdot Q_0\rangle_R & ||d_0\cdot Q_0||_R^2 \end{bmatrix} \tag{2.5}$$

and

$$A(U) = \begin{bmatrix} ||\mathcal{R}_U d_0||_R^2 & \langle \mathcal{R}_U d_0, \mathcal{R}_U d_0\cdot Q_0\rangle_R \\ \langle \mathcal{R}_U d_0, \mathcal{R}_U d_0\cdot Q_0\rangle_R & ||\mathcal{R}_U d_0\cdot Q_0||_R^2 \end{bmatrix}\ , \tag{2.6}$$

where $\mathcal{R}_U$ denotes the $H(R)$ orthogonal projector for the $H(R)$ subspace

$$R_U := \text{span}\{R(\cdot,u_i)|u_i\ \varepsilon\ U\}\quad .$$

Equations (2.4) - (2.6) have important implications for the optimal spacing problem. To illustrate this point consider the case of location parameter estimation when σ is assumed known. For a particular spacing, U, since $\sigma_B^2 A(U)^{-1}$ is the asymptotic variance-covariance matrix of the ABLUE it follows that

$$\mathrm{ARE}(\mu(U)) = ||\mathcal{R}_U d_0||_R^2/||d_0||_R^2$$
$$= ||d_0||_R^{-2}[||d_0||_R^2 - ||d_0 - \mathcal{R}_U d_0||_R^2]$$

as a result of the Pythagorean theorem. Thus, maximizing ARE$(\mu(U))$ with respect to U is equivalent to minimizing $||d_0 - \mathcal{R}_U d_0||_R^2$ over all $U \varepsilon D_n$. However, from (2.4),

$$||d_0 - \mathcal{R}_U d_0||_R^2 = ||d_0' - (\mathcal{R}_U d_0)'||_{L^2}^2$$
$$= ||d_0' - \mathcal{R}_U' d_0'||_{L^2}^2 \quad ,$$

where $\mathcal{R}_U'$ is the $L^2[0,1]$ orthogonal projection operator for the (L^2) subspace

$$R_U' = \mathrm{span}\{ \frac{\partial R(u,u_i)}{\partial u} \mid u_i \ \varepsilon \ U\} . \qquad (2.7)$$

Reference to (2.2) verifies that R_U' consists of splines of order 1 (piecewise constants)with knots or breakpoints at the elements of U. Therefore, the optimal spacing problem is now seen to coincide with the following variable knot spline approximation problem for d_0': find $U^* \varepsilon D_n$ such that

$$||d_0' - \mathcal{R}_{U^*}' d_0'||_{L^2} = \inf_{U \varepsilon D_n} ||d_0' - \mathcal{R}_U' d_0'||_{L^2} \quad . \qquad (2.8)$$

To ascertain how problem (2.8) relates to the usual type of variable knot piecewise constant approximation problem first note that for any $U \varepsilon D_n$ the elements in R_U' are orthogonal to the unit function, 1, in $L^2[0,1]$. Thus, the set of all splines of order 1 with knots at U, S_U, may be written

$$S_U = \text{span}\{1\} \oplus R_U' \ .$$

As $d_0 \varepsilon H(R)$ requires that $d_0(0)=d_0(1)=0$ it follows that $d_0' \perp$ span $\{1\}$, as well. Therefore, $d_0'-R_U'd_0' \perp S_U$ (in L^2) and, consequently, $R_U'd_0'$ is the best $L^2[0,1]$ approximation to d_0' from S_U. Now, let U* be defined as in (2.8) and let $\mathcal{S}_U$ denote the $L^2[0,1]$ orthogonal projector for S_U. Then $R_{U*}'d_0'$ satisfies

$$||d_0'-R_{U*}'d_0'||_{L^2} = \inf_{U\varepsilon D_n} ||d_0'-\mathcal{S}_U d_0'||_{L^2} \ . \qquad (2.9)$$

Equation (2.9) has the consequence that, for location parameter estimation, the optimal spacing problem is equivalent to finding the best set of knots for piecewise constant approximation of d_0' in $L^2[0,1]$. An analogous result holds for scale parameter estimation.

The preceding discussions are now summaried by way of the following theorem.

Theorem 1. If $d_0(d_0 \cdot Q_0)$ is in H(R) then the following three problems are equivalent:

(i) Optimal spacing selection for the ABLUE of $\mu(\sigma)$ when $\sigma(\mu)$ is known.

(ii) Minimum variance design selection for $\beta_1(\beta_2)$ when $\beta_2(\beta_1)$ is known in model (2.3).

(iii) Optimal knot selection for the best $L^2[0,1]$ approximation of $d_0'([d_0 \cdot Q_0]')$ by splines of order 1.

It is of interest to note the importance of Theorem 1 with regard to problems (ii) and (iii). From a regression design perspective the values of optimal spacings provided in references [1,3,4,5,10-14,17,18,25,28,29] may now be viewed as optimal designs for a regression problem with regression function $d_0(d_0 \cdot Q_0)$ and Brownian bridge covariance structure (these are also optimal designs for models having the Brownian motion covariance kernel, min(s,t), if a design point at 1 is appended) whereas from an approximation theory point of view they may be considered as optimal knot locations for piecewise constant approximation of $d_0'([d_0 \cdot Q_0]')$. *The optimal spacing literature, therefore, provides a readily available source for the optimal designs* (in the context of model (2.3)) *and optimal knots which correspond to a rich set of functions*. For this reason, it should be of considerable value, for comparison or other purposes, when alternative design or knot selection schemes are being considered.

3. *Asymptotic results*. In this section the asymptotic properties of certain spacing selection schemes will be analyzed. It will be seen that, in certain cases, it is possible to characterize the asymptotic behaviour of spacing sequences with regards to various criteria for measuring the size of A(U). In addition, spacing sequences that are asymptotically optimal (in a sense to be defined) for the optimality criteria $|A(U)|$, $V(\mu(U)) + V(\sigma(U))$ and ARE $(\mu(U))$ + ARE$(\sigma(U))$ will be provided. The elements of such sequences can be utilized to provide an approximate, easily computed, solution to the problems of optimal and suboptimal spacing selection.

For a nonnegative matrix B, let $\psi(B)$ denote either $|B|$ or trBM where M is a specified nonnegative matrix. Then the performance of a spacing sequence, $\{U_n\}_{n=1}^{\infty}$ $U_n \varepsilon D_n$, can be determined from a regret point of view by examining the asymptotic behaviour of $\psi(A)-\psi(A(U_n))$ or $\psi(A(U_n)^{-1}) - \psi(A^{-1})$. A sequence satisfying

$$\lim_{n\to\infty}[\inf_{U\varepsilon D_n} \psi(A(U)^{-1})-\psi(A^{-1})][\psi(A(U_n)^{-1})-\psi(A^{-1})]^{-1} = 1 \qquad (3.1)$$

is termed asymptotically ψ1-optimum whereas one satisfying

$$\lim_{n\to\infty}[\psi(A) - \sup_{U\varepsilon D_n} \psi(A(U))][\psi(A)-\psi(A(U_n))]^{-1} = 1 \qquad (3.2)$$

is said to be asymptotically ψ2-optimum (this terminology is due to Sacks and Ylvisaker (1968)). When only one parameter is unknown both (3.1) and (3.2) may be stated in terms of ARE's. If, for instance, only μ is unknown both (3.1) and (3.2) are equivalent to

$$\lim_{n\to\infty}[1- \sup_{U\varepsilon D_n} \text{ARE}(\mu(U))][1-\text{ARE}(\mu(U_n))]^{-1} = 1 \ . \qquad (3.3)$$

As in Eubank (1981), density functions will be utilized to generate spacing sequences. Let k be a continuous density on [0,1] with associated quantile function κ. Then k, or equivalently κ, defines a spacing sequence whose n-th element is $U_n=\{\kappa(\frac{1}{n+1}),\ldots,\kappa(\frac{n}{n+1})\}$. This sequence is called the regular sequence (RS) generated by k or κ and this relationship is indicated by $\{U_n\}$ is RS(κ). Since κ, or at least some of its values, must be known in order to obtain the U_n most of the results which follow will be stated in terms of κ rather than k (to translate conditions on κ to conditions on k it is only necessary to employ the change of variable $x = \kappa(u)$ and the relation $\kappa'(u) = 1/k(\kappa(u))$). In the context of knot (design) selection, such k's are frequently termed knot density functions and κ is referred to as a knot quantile function.

Let $W^{2,2}(a,b)$ denote the Sobolev space of functions on (a,b) possessing a second distribution derivative in $L^2[a,b]$ and define the space $W^{2,2}_{loc}(0,1)$ as the set of all functions in $W^{2,2}(a,b)$ if $0 < a < b < 1$. The next theorem details the asymptotic behaviour of spacing sequences for the trace and determinant criteria. The primary emphasis is upon spacings generated by a knot or spacing quantile function. In particular, knot density functions which generate asymptotically optimal spacing sequences are provided.

<u>Theorem 2.</u> Let $\kappa \in C[0,1]$ be a knot quantile function with a bounded piecewise continuous derivative κ' having the property that the set of all points where κ' is zero or discontinuous has content zero and has neither 0 or 1 as accumulation points. Using g to denote either d_0 or $d_0 \cdot Q_0$, it is assumed that $g \in W^{2,2}_{loc}(0,1) \cap H(R)$ and that for each function there is a corresponding $\delta > 0$ and a monotone function h on $I=(0,\delta] \cup [1-\delta,1]$ which satisfies

$$h(x) \geq |g''(x)| \quad \text{for all } x \in I, \tag{3.4}$$

and

$$\int_I |h(\kappa(x))|^2 \kappa'(x)^3 dx < \infty . \tag{3.5}$$

Under these hypotheses the following results hold.

(i) Let $\psi(B) = \mathrm{tr}BM$ and define

$$\phi(u) := (d_0''(u), (d_0 \cdot Q_0)''(u))^t. \tag{3.6}$$

Then, if $\{U_n\}$ is RS(κ)

$$\lim n^2\{\mathrm{tr}AM - \mathrm{tr}A(U_n)M\} = \frac{1}{12}\int_0^1 [\phi(\kappa(x))^t M\phi(\kappa(x))]\kappa'(x)^3 dx. \tag{3.7}$$

If d_0 and $d_0 \cdot Q_0$ are in $C^2[0,1] \cap H(R)$ then the RS generated by the density

$$k^*(x) = [\phi(x)^t M\phi(x)]^{1/3} \Big/ \{\int_0^1 [\phi(s)^t M\phi(s)]^{1/3} ds\} \tag{3.8}$$

is asymptotically $\psi 2$-optimum. In addition, if M is positive, the RS generated by the density

$$k^*(x) = [\phi(x)^t A^{-1} M A^{-1}\phi(x)]^{1/3} \Big/ \{\int_0^1 [\phi(s)^t A^{-1} M A^{-1}\phi(s)]^{1/3} ds\} \tag{3.9}$$

is asymptotically $\psi 1$-optimum.

(ii) Let $\psi(B) = |B|$. Then, if $\{U_n\}$ is RS(κ)

$$\lim_{n\to\infty} n^2 |A|\{1-\mathrm{ARE}(\mu(U_n),\sigma(U_n))\} = \frac{1}{12}\int_0^1 [\phi(\kappa(x))^t A^{-1}\phi(\kappa(x))]\kappa'(x)^3 dx. \tag{3.10}$$

If both d_0 and $d_0 \cdot Q_0$ are in $C^2[0,1] \cap H(R)$ then a RS which is asymptotically $\psi 1$ and $\psi 2$ optimum is generated by the density

$$k^*(x) = [\phi(x)^t A^{-1}\phi(x)]^{1/3} \Big/ \{\int_0^1 [\phi(s)^t A^{-1}\phi(s)]^{1/3} ds\} . \tag{3.11}$$

Proof. The asymptotic optimality of the sequences generated by (3.8), (3.9), and (3.11) is an immediate consequence of results given in Sacks and Ylvisaker (1968). To obtain (3.7) and (3.10) first note that under the present assumptions the work of Pence and Smith (1981) (or Barrow and Smith (1978) under stronger conditions on κ) in conjunction with Theorem 1 has the consequence that

$$\lim_{n\to\infty} n^2||g-\mathcal{R}_{U_n} g||_R^2 = \frac{1}{12}\int_0^1 [g''(\kappa(x))]^2 \kappa'(x)^3 dx, \tag{3.12}$$

where g has been utilized to denote either d_0 or $d_0 \cdot \mathcal{Q}_0$. Equation (3.12), along with the technique utilized in proving Theorems 4.1 and 4.2 in Sacks and Ylvisaker (1968), then gives the desired results.

Theorem 2 provides the necessary tools for analyzing the asymptotic behaviour of the various spacing selection schemes. Using (3.11) in (3.10) and the asymptotic optimality of the corresponding RS it follows that for spacings obtained by maximizing Ogawa's ARE expression

$$\lim_{n\to\infty} n^2|A|\{1-\sup_{U\varepsilon D_n} ARE(\mu(U),\sigma(U))\} = \frac{1}{12}\{\int_0^1 [\phi(x)^t A^{-1}\phi(x)]^{1/3} dx\}^3 \tag{3.13}$$

This is to be compared to the approach of maximizing the sum of the ARE's utilized by Hassanein (1977). In this latter case, using (3.7) and (3.8) with $M^{-1} = \mathrm{diag}(||d_0||_R^2, ||d_0 \cdot \mathcal{Q}_0||_R^2)$ one obtains

$$\lim_{n\to\infty} n^2\{trAM-\sup_{U\varepsilon D_n} trA(U)M\} = \frac{1}{12}\left\{\int_0^1 \left[\left(\frac{d_0''(x)}{||d_0||_R}\right)^2 + \left(\frac{(d_0 \cdot \mathcal{Q}_0)''(x)}{||d_0 \cdot \mathcal{Q}_0||_R}\right)^2\right]^{1/3} dx\right\}^3 . \tag{3.14}$$

Finally, if the criteria is minimization of the sum of variances, as in Hassanein (1969a, 1969b) and Eisenberger and Posner (1965), then from (3.9) and Theorem 4.5 of Sacks and Ylvisaker (1968) with M = I, the limiting behaviour is given by

$$\lim_{n\to\infty} n^2\{\inf_{U\varepsilon D_n} trA(U)^{-1}-trA^{-1}\} = \frac{1}{12}\{\int_0^1 [\phi(x)^t A^{-2}\phi(x)]^{1/3} dx\}^3 . \tag{3.15}$$

Thus, for symmetric distributions, where A is a diagonal matrix, spacings which are D-optimal or maximize $ARE(\mu(U)) + ARE(\sigma(U))$ will have the same asymptotic behaviour. However, the asymptotic properties of these spacings will, in general, differ from that of spacings obtained by minimizing the sum of the variances, even for symmetric distribution. All three spacing selection schemes will behave similarly for distributions such as the Cauchy where A is a constant multiple of the identity. In fact, for the Cauchy distribution asymptotically optimal spacing sequences for minimization of $|A(U)|$, $V(\mu(U)) + V(\sigma(U))$, and $ARE(\mu(U)) + ARE(\sigma(U))$ are all generated by the same knot quantile function, $\kappa^*(x)=x$.

The density (3.11) was also derived for use in spacing selection by Sarndal (1962) using variational methods and under more stringent conditions. For an alternative approach see Eubank (1981).

4. Robust estimation. Chan and Rhodin (1980) have proposed a technique which utilizes the ABLUE to accomplish the robust estimation of the location parameter of a symmetric distribution. They assume that the true underlying distribution is a member of the Tukey's lambda family or is a normal, double exponential or Cauchy distribution (or, at least, is well modelled by one or more of these laws). In addition, they assume that F_0 belongs to a known finite subset, $\mathcal{L}$, of these families of distributions. Let $ARE(\mu(U)|G)$ denote the ARE corresponding to the spacing U when estimation is to be accomplished under the assumption that the data has the distribution $G \varepsilon \mathcal{L}$. Then, to estimate μ, Chan and Rhodin take as their guess for F_0 any distribution $F^* \varepsilon \mathcal{L}$ which satisfies

$$\min_{G\varepsilon\mathcal{L}} ARE(\mu(U(F^*))|G) = \max_{F\varepsilon\mathcal{L}} \min_{G\varepsilon\mathcal{L}} ARE(\mu(U(F))|G), \tag{4.1}$$

where U(F) is an optimal spacing for the distribution F. This approach requires that the function $\mathrm{ARE}(\mu(U(F))|G)$ must be tabulated for all pairwise combinations of laws in L and for each value of n that is to be considered (Chan and Rhodin provide tables for n = 2(1)5).

From the results in Section 2 it follows than the procedure utilized by Chan and Rhodin (1980) is equivalent to: select an $F^* \varepsilon L$ such that

$$\max_{G\varepsilon L} ||d_G - R_{U(F^*)} d_G||_R^2 = \min_{F\varepsilon L} \max_{G\varepsilon L} ||d_G - R_{U(F)} d_G||_R^2 , \qquad (4.2)$$

where d_G is the density-quantile function for $G\varepsilon L$. This suggests the following asymptotic approach to robust spacing selection. Under conditions such as those in Theorem 3.1, define for each $F\varepsilon L$ the knot density function

$$k_F(x) = \{d_F''(x)\}^{2/3} / \int_0^1 \{d_F''(s)\}^{2/3} ds \qquad (4.3)$$

with corresponding knot quantile function denoted κ_F. If $d_0 \ \varepsilon \ C^2[0,1]$ the density k_F generates a sequence of asymptotically optimal (in the sense of (3.3)) spacings, $\{U_n(F)\}$, for location parameter estimation when F is the true parent distribution of the data (c.f. Eubank (1981)). Then, from (3.12)

$$||d_G - R_{U_n(F)} d_G||_R^2 = \frac{1}{12n^2} \int_0^1 [d_G''(\kappa_F(x))]^2 \kappa_F'(x)^3 dx + o(n^{-2}). \qquad (4.4)$$

Hence, an asymptotic version of (4.1) is: select $F^* \varepsilon L$ so that

$$\max_{G\varepsilon L} \int_0^1 [d_G''(\kappa_{F^*}(x))]^2 \kappa_{F^*}'(x)^3 dx = \min_{F\varepsilon L} \max_{G\varepsilon L} \int_0^1 [d_G''(\kappa_F(x))]^2 \kappa_F'(x)^3 dx. \qquad (4.5)$$

To determine F* for a given L one must evaluate (usually by numerical techniques) the function $\int_0^1 [d_G''(\kappa_F(x))]^2 \kappa_F'(x)^3 dx$ for all pairwise combinations of laws in L. However, in contrast to the procedure suggested by Chan and Rhodin, the resulting tabulation suffices for all values of

n. An estimator of μ based on n quantiles is then provided through use of the spacing $U_n(F^*)$ and the corresponding coefficients for F* that may be obtained from Ogawa (1951).

The solution (4.5) is applicable to any (finite) set of laws $\mathcal{L}$, whether symmetric or not, provided conditions such as those in Theorem 2 are satisfied by the elements of $\mathcal{L}$. A scale parameter version of (4.5) can be obtained by using $d \cdot Q$ rather than d, in (4.3) and (4.5). For the simultaneous estimation of μ and σ either (3.6) and (3.8) or (3.10) and (3.11) can be utilized to construct an analogous criterion for robust spacing selection.

Acknowledgement

The author would like to express his gratitude to Professors Pence and Smith for providing a preprint of their manuscript and to Professor Smith for several helpful conversations during the course of this research.

Randall L. Eubank
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

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.

Barrow, D.L. and Smith, P.W. (1978). Asymptotic properties of best $L^2[0,1]$ approximation by splines with variable knots. *Quarterly of Applied Math 36*, 293-304.

Bloch, D. (1966). A note on the estimation of the location parameter of the Cauchy distribution. *J. Amer. Statist. Assoc. 61*, 852-855.

Chan, L.K. (1970). Linear estimation of the location and scale parameter of the Cauchy distribution based on sample quantiles. *J. Amer. Statist. Assoc. 65*, 851-859.

Chan, L.K. and Rhodin, L.S. (1980). Robust estimation of location using optimally chosen sample quantiles. *Technometrics 22*, 225-237.

Chernoff, H. (1971). A note on optimal spacings for systematic statistics. Stanford Univ. Tech. Rep. No. 70.

Eisenberger, I. and Posner, E.C. (1965). Systematic statistics used for data compression in space telemetry. *J. Amer. Statist. Assoc. 60*, 97-133.

Eubank, R.L. (1979). A density-quantile function approach to the selection of order statistics for location and scale parameter estimation. Texas A & M University Tech. Rep. No. A10.

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

Gupta, S.S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika 53*, 565-570.

Hassanein, K.M. (1968). Analysis of extreme value data by sample quantiles. *J. Amer. Statist. Assoc. 63*, 877-888.

Hassanein, K.M. (1969a). Estimation of the parameters of the extreme value distribution by use of two or three order statistics. *Biometrika 56*, 429-436.

Hassanein, K.M. (1969b). Estimation of the parameters of the logistic distribution by sample quantiles. *Biometrika 56*, 684-687.

Hassanein, K.M. (1971). Percentile estimators for the parameters of the Weibull distribution. *Biometrika 58*, 673-676.

Hassanein, K.M. (1972). Simultaneous estimation of the parameters of the extreme value distribution by sample quantiles. *Technometrics 14*, 63-70.

Hassanein, K.M. (1977). Simultaneous estimation of the location and scale parameter of the gamma distribution by linear functions of order statistics. *Skand. Aktuarietidskr. 56*, 120-128.

Kulldorf, G. (1963). On the optimum spacing of sample quantiles from a normal distribution. Part 1. *Skand. Aktuarietidskr 46*, 143-146.

Kulldorf, G. and Vännman, K. (1973). Estimation of the location and scale parameter of the Pareto distribution by linear functions of order statistics. *J. Amer. Statist. Assoc. 68*, 218-227.

Mosteller, F. (1946). On some useful inefficient statistics. *Ann. Math. Statist. 17*, 175-213.

Ogawa, J. (1951). Contributions to the theory of systematic statistics, I. *Osaka Math. J. 3*, 131-142.

Parzen, E. (1961a). An approach to time series analysis. *Ann. Math. Statist. 32*, 951-989.

Parzen, E. (1961b). Regression analysis of continuous parameter time series. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I.*, 469-489.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc. 74*, 105-121.

Pence, D.D. and Smith, P.W. (1981). Asymptotic properties of best $L_p[0,1]$ approximation by splines. *SIAM J. Math. Anal.*, to appear.

Rhodin, L.S. (1976). Optimum spacing for the ABLE of location or scale parameters in Tukey's lambda family. Statistical Research Rep. No. 1976-10. Institute of Mathematics and Statistics, University of Umea, Sweden.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist. 37*, 66-89.

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. (1958). Estimation problems in the exponential distribution using order statistics. *Proceedings of the Statistical Techniques in Missile Evaluation Symposium*, Blacksburg, Va., 123-173.

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

Särndal, C. (1962). *Information from Censored Samples.* Stockholm: Almqvist and Wiksell.

Särndal, C. (1964). Estimation of the parameters of the gamma distribution by sample quantiles. *Technometrics 6*, 405-414.