LOCATION AND SCALE PARAMETER ESTIMATION FROM RANDOMLY CENSORED DATA

by

R. L. Eubank and V. N. La Riccia

Technical Report No. 166
Department of Statistics ONR Contract

August, 1982

Research sponsored by the Office of Naval Research
Contract N00014-82-k-0207

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, TX 75275

LOCATION AND SCALE PARAMETER ESTIMATION FROM RANDOMLY CENSORED DATA

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

V. N. LaRiccia
Department of Math. Sciences
University of Delaware
Newark, Delaware 19711

Key Words & Phrases: quantile function; estimation; location and scale parameters; random censoring; reproducing kernel Hilbert space.

ABSTRACT

The problem of location and scale parameter estimation from randomly censored data is analyzed through use of a regression model for the Kaplan-Meier quantile process. Continuous time regression techniques are employed to construct estimators that are both asymptotically normal and efficient. Estimators with a particularly simple form are obtained for the Koziol-Green model for random censorship. In the event of no censoring the regression model, and resulting estimators, reduce to those proposed by Parzen (1979 a, b).

1. INTRODUCTION

In this paper the problem of location and scale parameter estimation from randomly censored data is formulated as a regression analysis problem for the Kaplan-Meier quantile process. Under certain regularity conditions, estimators are derived that

are both asymptotically normal and efficient. In addition, this approach to the problem is seen, through examples and discussion, to provide a framework for estimation that can be used to motivate and generalize estimation procedures suggested by other authors for both censored and uncensored samples.

Let $T_1,\ldots,T_n$ denote the true survival times of n individuals which are assumed to be i.i.d. random variables having common distribution function (d.f.)

$$F(t) = F_0(\frac{t-\mu}{\sigma}) \tag{1.1}$$

where μ and σ are, respectively, unknown location and scale parameters and F_0 is a known distributional form. Further, let $Y_1,\ldots,Y_n$ denote n i.i.d. censoring random variables with common d.f. H that are independent of the T_i's. In the random censoring model one observes not the T_i's but, instead, the pairs of random variables (Z_i,δ_i) where $Z_i = \min(T_i,Y_i)$, $\delta_i = \chi_{\{T_i \le Y_i\}}$ and χ denotes the indicator function. The d.f. of the Z_i's, F^*, is then given by the relation

$$1 - F^*(z) = [1-F_0(\frac{z-\mu}{\sigma})][1-H(z)] \ . \tag{1.2}$$

An important problem associated with this model is the estimation of the parameters μ and σ from the observed data.

Let $F_n(t)$ denote the Kaplan-Meier estimator of the d.f. F(t) (Kaplan and Meier (1958)) with associated empirical quantile function defined by

$$Q_n(u) = \inf\{t: F_n(t) \ge u\} \ . \tag{1.3}$$

Functionals of F_n, and Q_n as well as the Bayesian generalization of F_n have been used in the nonparametric estimation of various types of parameters, such as the mean survival time, by Breslow and Crowley (1974), Sander (1975a, 1975b, 1975c), Susarla and Van Ryzin (1980) and Reid (1981). In contrast, we develop estimators that are functionals of Q_n but are applicable to general μ and σ under the assumption that the precise form of F_0 is known. The model that is assumed is, therefore, the censored sample analog

of the classical location and scale parameter model. The proposed estimators are seen to have a closed, computationally simple, form which should provide savings over other techniques, such as maximum likelihood, which are frequently used in this situation (c.f. Kalbfleisch and Prentice (1980)). For the estimation of parameters other than those of the location/scale variety the reader is referred to Eubank and LaRiccia (1981) for a minimum distance approach.

In Section 2 we present our principal result with the proof provided in Section 3. Our approach is based on a continuous time regression model for the quantile process, $Q_n(\cdot)$, which allows us to view the problem of estimating μ and σ from a regression analysis perspective. In the important case of the Koziol-Green model for random censorship this approach results in estimators of μ and σ having a simple closed form. Closed form estimators are also provided for the general model (1.2) that are seen to be asymptotically normal and efficient. The relationships between these estimators and estimators proposed in Eubank and LaRiccia (1982), LaRiccia (1982), Parzen (1979a,b) and Weiss (1964) and Weiss and Wolfowitz (1970) are also explored.

2. ESTIMATOR DERIVATION AND RESULTS

In this section estimators of μ and σ are presented that are functionals of the empirical quantile function Q_n. The techniques used in the construction of these estimators are suggested by the approach of Parzen (1979a) who, in the uncensored case, showed that location and scale parameter estimation could be viewed as a continuous time regression problem for the quantile process. We begin by showing that a similar result holds for randomly censored data. First, however, some notational preliminaries are required.

Assume that F_0 admits a continuous density $f_0 = F_0'$ and define the quantile function Q_0 by

$$Q_0(u) = \inf\{t: F_0(t) \geq u\} .$$

The density-quantile function corresponding to F_0 is then given by $f_0Q_0(u) = f_0(Q_0(u))$, $0 \leq u \leq 1$. If F_0 is strictly increasing the quantile functions for F and F_0 are related by

$$Q(u) = \mu + \sigma Q_0(u) \tag{2.1}$$

whereas their density-quantile functions satisfy

$$fQ(u) = \frac{1}{\sigma} f_0Q_0(u) . \tag{2.2}$$

For a proof of (2.1) as well as a discussion of quantile and density-quantile functions and their properties the reader is referred to Parzen (1979a).

Let β_1, β_2 be two constants satisfying $0 < \beta_1 < \beta_2 < 1$. The work of Sander (1975a) may be used to show that, on the interval $[\beta_1,\beta_2]$, $\sqrt{n}\, fQ(u)\{Q_n(u)-Q(u)\}$ converges in distribution to a zero mean Gaussian process $\{X(u),\ u \in [\beta_1,\beta_2]\}$ with covariance kernel

$$K(s,t) = (1-t)G(s), \quad s < t , \tag{2.3}$$

where

$$G(s) = (1-s)\int_0^s \frac{1}{(1-u)^2[1-HQ(u)]}\, du . \tag{2.4}$$

and $HQ(u) = H(Q(u))$. Using (2.1) and (2.2) this is seen to have the implication that, asymptotically, location and scale parameter estimation can be formulated as a continuous time regression problem for the quantile process via the model

$$f_0Q_0(u)Q_n(u) = \mu f_0Q_0(u) + \sigma f_0Q_0(u)Q_0(u) + \sigma_c X(u),\ u\in[\beta_1,\beta_2], \tag{2.5}$$

where $\sigma_c = \sigma/\sqrt{n}$. Using the work of Csörgő and Révész (1978) a similar model was developed by Parzen (1979a), for the uncensored case, wherein X is a Brownian bridge process.

Model (2.5) will be used in conjunction with the theory of time series analysis by reproducing kernel Hilbert space (RKHS) methods to devise estimators for μ and σ. The first step in this procedure is to determine the form of $\mathcal{H}(K)$, the RKHS generated by the covariance kernel (2.3). It is well known that $\mathcal{H}(K)$ is unique and is congruent to the X process (the reader is referred to

Aronszajn (1950) and Parzen (1961a, 1961b) for a more detailed presentation of respectively, the theory of reproducing kernels and their role in inference for stochastic processes). The congruence between these spaces will allow us to obtain explicit estimator formulae through use of the $H(K)$ norm and inner product.

Using results given in Sacks and Ylvisaker (1966), $H(K)$ is found to consist of L^2-differentiable functions with the inner product of two functions m, $g \varepsilon H(K)$ given by

$$<g,m>_K = \int_{\beta_1}^{\beta_2} \left[\frac{g(u)}{1-u}\right]' \left[\frac{m(u)}{1-u}\right]' (1-u)^2 [1-HQ(u)]du + \frac{g(\beta_1)m(\beta_1)}{K(\beta_1,\beta_1)} . \tag{2.6}$$

The norm for $H(K)$ will be denoted by $||\cdot||_K$. Two alternative forms of (2.6) will also be useful in subsequent work. If H admits a continuous density, h, the inner product can be rewritten

$$\begin{aligned} <g,m>_K = & \int_{\beta_1}^{\beta_2} g'(u)m'(u)[1-HQ(u)]du \\ & + \sigma\int_{\beta_1}^{\beta_2} \left[\frac{g(u)}{1-u}\right]\left[\frac{m(u)}{1-u}\right](1-u)\frac{hQ(u)}{f_0Q_0(u)}du \\ & + g(\beta_1)m(\beta_1)\left[\frac{1}{K(\beta_1,\beta_1)} - \frac{1-HQ(\beta_1)}{1-\beta_1}\right] \\ & + \frac{g(\beta_2)m(\beta_2)}{1-\beta_2}[1-HQ(\beta_2)] \end{aligned} \tag{2.7}$$

and, if g is assumed to be twice continuously differentiable on $[\beta_1,\beta_2]$, another expression for (2.6) is furnished by

$$\begin{aligned} <g,m>_K = & -\int_{\beta_1}^{\beta_2}\{g''(u)+G''(u)[g(u)+g'(u)(1-u)](1-HQ(u))\}[1-HQ(u)]m(u)du \\ & + \frac{g(\beta_1)G'(\beta_1)-G(\beta_1)g'(\beta_1)}{G(\beta_1)}[1-HQ(\beta_1)]m(\beta_1) \\ & + \frac{g(\beta_2) + (1-\beta_2)g'(\beta_2)}{1-\beta_2}[1-HQ(\beta_2)]m(\beta_2) . \end{aligned} \tag{2.8}$$

Before proceeding further it should be noted that the inner product in $H(K)$ provides a natural measure of information. Thus, now, and in subsequent discussions, we impose the regularity condition that f_0Q_0 and the product of f_0Q_0 and Q_0, denoted $f_0Q_0 \cdot Q_0$, are elements of $H(K)$ and define the <u>information matrix</u>

$$I(\beta_1,\beta_2) = \begin{bmatrix} ||f_0Q_0||_K^2 & <f_0Q_0,\ f_0Q_0\cdot Q_0>_K \\ <f_0Q_0, f_0Q_0\cdot Q_0>_K & ||f_0Q_0\cdot Q_0||_K^2 \end{bmatrix} . \tag{2.9}$$

To justify this title for $I(\beta_1,\beta_2)$ consider the case of location parameter estimation. Noting that the densities for the uncensored and censored observations are, respectively,

$$f^u(x) = [1-H(x)]f(x) = [1-H(x)]\frac{1}{\sigma}f_0(\frac{x-\mu}{\sigma})$$

and

$$f^c(x) = [1-F(x)]h(x) = [1 - F_0(\frac{x-\mu}{\sigma})]h(x)$$

(c.f. Reid (1981)) the change of variable $x = Q(u)$ in (2.7) gives for instance, that when $\beta_1=0$, $\beta_2 = 1$

$$||f_0Q_0||_K^2 = \int_0^\infty \left[\frac{f_0'(\frac{x-\mu}{\sigma})}{f_0(\frac{x-\mu}{\sigma})}\right]^2 [1-H(x)]f_0(\frac{x-\mu}{\sigma})dx$$

$$+ \int_0^\infty \left[\frac{f_0(\frac{x-\mu}{\sigma})}{1-F_0(\frac{x-\mu}{\sigma})}\right]^2 [1-F_0(\frac{x-\mu}{\sigma})]h(x)dx$$

$$= \sigma^2 \left\{ \int_0^\infty \left[\frac{f'(x)}{f(x)}\right]^2 f^u(x)dx + \int_0^\infty \left[\frac{f(x)}{1-F(x)}\right]^2 f^c(x)dx \right\} .$$

By examination of the form of the likelihood for the observations, L, given, for example, in Kalbfleisch and Prentice (1980) and conditioning on the values of δ_i it can be verified that $\sigma_c^2\ ||f_0Q_0||_K^2 = E[(\partial \log L/\partial\mu)^2]$ and, as a result, is the usual Fisher information measure for location estimation. Similarly $\sigma_c^2\ ||f_0Q_0||_K^2$ measures information in the Fisher sense for the case we consider of $0 < \beta_1 < \beta_2 < 1$. Identical statements are readily seen to hold for the other elements of $I(\beta_1,\beta_2)$ as well.

The random variables congruent to f_0Q_0 and $f_0Q_0 \cdot Q_0$ are the functionals $\langle f_0Q_0, f_0Q_0 \cdot Q_n \rangle_K$ and $\langle f_0Q_0 \cdot Q_0, f_0Q_0 \cdot Q_n \rangle_K$. From model (2.5) and the theory of continuous time regression it now follows that an asymptotically normal and efficient estimator of $(\mu,\sigma)^t$ is

$$\begin{bmatrix} \mu^* \\ \sigma^* \end{bmatrix} = I^{-1}(\beta_1,\beta_2) \begin{bmatrix} \langle f_0Q_0, & f_0Q_0 \cdot Q_n \rangle_K \\ \langle f_0Q_0 \cdot Q_0, & f_0Q_0 \cdot Q_n \rangle_K \end{bmatrix} \tag{2.10}$$

which has variance-covariance matrix $\sigma_c^2 I^{-1}(\beta_1,\beta_2)$. Under the Koziol-Green model for random censoring (2.10) may be used directly to provide estimators of μ and σ as we now illustrate.

<u>Example 1.</u> Koziol and Green (1976) consider testing the hypothesis

$$H_0\colon F(t) = F_0(t;\underline{\theta}_0)$$

where $\underline{\theta}_0$ is a vector of <u>specified</u> parameter values that completely determine F_0. They assume that H and F are related by

$$1-H(t) = [1-F(t)]^{\alpha}$$

where α is termed the censoring parameter since it can be shown that $\nu = (1+\alpha)^{-1}$ is the expected proportion of uncensored observations. They show through an example that their model is useful for testing H_0. Unfortunately, the asymptotic distribution of their statistic depends on α which may be unknown. Csörgő and Horváth (1981) have proposed an alternative statistic that alleviates this difficulty by incorporating a strongly consistent estimator of ν, namely $\nu_n = n^{-1}\Sigma_{j=1}^n \delta_j$. See also Hollander and Proschan (1979) for other work in this area.

We now examine the form of μ^* and σ^* under the Koziol-Green model for (1.2), i.e. we assume that

$$1-H(t) = [1-F_0(\tfrac{t-\mu}{\sigma})]^{\alpha} . \tag{2.12}$$

Using form (2.6) of the inner product, it is readily verified that

$$\langle g,m \rangle_K = \int_{\beta_1}^{\beta_2} \left[\frac{g(u)}{1-u}\right]' \left[\frac{m(u)}{1-u}\right]' (1-u)^{2+\alpha} du + \frac{(\alpha+1)g(\beta_1)m(\beta_1)}{[(1-\beta_1)^{-\alpha+1}-(1-\beta_1)^2]} . \tag{2.13}$$

Then, from (2.8), we have

$$<f_0Q_0, f_0Q_0\cdot Q_n>_K = \int_{\beta_1}^{\beta_2} W^*_\mu(u)Q_n(u)du + \sum_{i=1}^{2} W^*_{\mu,\beta_i} Q_n(\beta_i) \qquad (2.14)$$

$$<f_0Q_0\cdot Q_0, f_0Q_0\cdot Q_n>_K = \int_{\beta_1}^{\beta_2} W^*_\sigma(u)Q_n(u)du + \sum_{i=1}^{2} W^*_{\sigma,\beta_i} Q_n(\beta_i) \qquad (2.15)$$

where

$$W^*_\mu(u) = -\{(f_0Q_0)''(u)-\alpha\left[\frac{f_0Q_0(u)}{1-u}\right]'\}(1-u)^\alpha f_0Q_0(u) \ , \qquad (2.16)$$

$$W^*_{\mu,\beta_1} = \{f_0Q_0(\beta_1)\ \frac{[\alpha+(1-\beta_1)^{\alpha+1}]}{(1-\beta_1) - (1-\beta_1)^{\alpha+2}}$$

$$- (f_0Q_0)'(\beta_1)\}(1-\beta_1)^\alpha f_0Q_0(\beta_1) \ , \qquad (2.17)$$

$$W^*_{\mu,\beta_2} = [\ \frac{1}{1-\beta_2}\ f_0Q_0(\beta_2) + (f_0Q_0)'(\beta_2)](1-\beta_2)^\alpha f_0Q_0(\beta_2), \qquad (2.18)$$

and W^*_σ, W^*_{σ,β_1} and W^*_{σ,β_2} are defined similarly by replacing f_0Q_0 by the product $f_0Q_0\cdot Q_0$ in (2.16) - (2.18). Thus, if α is known, (2.10) may be used directly to estimate μ and σ. If α is unknown a strongly consistent estimator of α is

$$\alpha_n = \nu_n^{-1}(1-\nu_n)$$

that may be used in the weights without influencing the asymptotic properties of the estimators.

The estimators μ^* and σ^* could also be used in conjunction with the test statistics of Koziol and Green and Csörgő and Horváth to test hypotheses of the form H_0: $F(x) = F_0(\frac{x-\mu}{\sigma})$ where the parameter values are, in this case, unknown. We have not investigated the properties of such a test but note that if one utilizes a smoothed (eg., piecewise linear) version of Q_n our regression formulation suggests that other useful measures of goodness-of-fit might be based on $||f_0Q_0(\cdot)\{Q_n(\cdot)-\mu^*-\sigma^*Q_0(\cdot)\}||_K^2$.

In general, the inner product $<\cdot,\cdot>_K$ will involve the unknown parameters due to the term $1-HQ(u) = 1-H(\mu+\sigma Q_0(u))$ appearing in (2.4), (2.6) and (2.8). As a result, $(\mu^*,\sigma^*)^t$ will

frequently not be computable. We, therefore, proceed to derive estimators that are always computable, when H is known, by appropriately estimating the unknown quantities in (2.10). The case of H unknown is discussed in Remark 1 below.

The difficulties encountered with $(\mu^*,\sigma^*)^t$ can be circumvented by replacing Q, where it appears, by Q_n in our previous formulation. Therefore, let

$$<g,m>_{\hat{K}}=\int_{\beta_1}^{\beta_2}\left[\frac{g(u)}{1-u}\right]'\left[\frac{m(u)}{1-u}\right]'(1-u)^2[1-HQ_n(u)]du+\frac{g(\beta_1)m(\beta_1)}{\hat{K}(\beta_1,\beta_1)} \quad (2.19)$$

where $\hat{K}(\beta_1,\beta_1) = (1-\beta_1)\hat{G}(\beta_1)$ and

$$\hat{G}(\beta_1) = (1-\beta_1)\int_0^{\beta_1}\frac{1}{(1-u)^2[1-HQ_n(u)]}du\ . \quad (2.20)$$

An estimator of $I(\beta_1,\beta_2)$ is then provided by

$$\hat{I}(\beta_1,\beta_2) = \begin{bmatrix} ||f_0Q_0||^2_{\hat{K}} & <f_0Q_0,f_0Q_0\cdot Q_0>_{\hat{K}} \\ <f_0Q_0,f_0Q_0\cdot Q_0>_{\hat{K}} & ||f_0Q_0\cdot Q_0||^2_{\hat{K}} \end{bmatrix}. \quad (2.21)$$

To obtain estimators of the other quantities in (2.10), assume that both f_0Q_0 and $f_0Q_0\cdot Q_0$ are twice continuously differentiable and replace expressions of the form (2.22)

$$A(u;g) = G''(u)[g(u)+g'(u)(1-u)](1-HQ(u))^2=\sigma(1-u)\left[\frac{g(u)}{1-u}\right]'\frac{hQ(u)}{f_0Q_0(u)}$$

and

$$B(u) = \frac{G'(\beta_1)}{G(\beta_1)}[1-HQ(\beta_1)] = \frac{1}{1-\beta_1}\left\{\frac{1}{G(\beta_1)} - [1-HQ(\beta_1)]\right\}, \quad (2.23)$$

which appear in (2.8), by

$$A_n(u;g) = \hat{\sigma}(1-u)\left[\frac{g(u)}{1-u}\right]'\left[\frac{hQ_n(u)}{f_0Q_0(u)}\right] \quad (2.24)$$

and

$$B_n(\beta_1) = \frac{1}{1-\beta_1}\left\{\frac{1}{\hat{G}(\beta_1)} - [1-HQ_n(\beta_1)]\right\}, \quad (2.25)$$

where $\hat{\sigma}$ is any estimator converging in probability to σ (denoted $\hat{\sigma} \to_p \sigma$). Of course if σ is known, and only μ is being estimate, one would use its value in (2.24) and dispense with the auxiliary estimator $\hat{\sigma}$. Now define the location weights

$$W_\mu(u) = -\{(f_0Q_0)''(u)[1-HQ_n(u)]-A_n(u;f_0Q_0)\}f_0Q_0(u) \tag{2.26}$$

$$W_{\mu,\beta_1} = \{f_0Q_0(\beta_1)B_n(\beta_1) - (f_0Q_0)'(\beta_1)[1-HQ_n(\beta_1)]\}f_0Q_0(\beta_1) \tag{2.27}$$

$$W_{\mu,\beta_2} = \{\frac{1}{1-\beta_2} f_0Q_0(\beta_2)+(f_0Q_0)'(\beta_2)\}[1-HQ_n(\beta_2)]f_0Q_0(\beta_2) \tag{2.28}$$

and the scale weights

$$W_\sigma(u) = -\{(f_0Q_0\cdot Q_0)''(u)[1-HQ_n(u)]-A_n(u;f_0Q_0\cdot Q_0)\}f_0Q_0(u) \tag{2.29}$$

$$W_{\sigma,\beta_1} = \{f_0Q_0(\beta_1)Q_0(\beta_1)B_n(\beta_1)-(f_0Q_0\cdot Q_0)'(\beta_1)[1-HQ_n(\beta_1)]\}f_0Q_0(\beta_1) \tag{2.30}$$

$$W_{\sigma,\beta_2} = \{\frac{1}{1-\beta_2}f_0Q_0(\beta_2)Q_0(\beta_2)+(f_0Q_0\cdot Q_0)'(\beta_2)\}[1-HQ_n(\beta_2)]f_0Q_0(\beta_2). \tag{2.31}$$

Estimators for the functional $\langle f_0Q_0, f_0Q_0\cdot Q_n\rangle_K$ and $\langle f_0Q_0\cdot Q_0, f_0Q_0\cdot Q_n\rangle_K$ can then be expressed as

$$\langle f_0Q_0, f_0Q_0\cdot Q_n\rangle_{\hat{K}} = \int_{\beta_1}^{\beta_2} W_\mu(u)Q_n(u)du + \sum_{i=1}^{2} W_{\mu,\beta_i}Q_n(\beta_i) \tag{2.32}$$

and

$$\langle f_0Q_0\cdot Q_0, f_0Q_0\cdot Q_n\rangle_{\hat{K}} = \int_{\beta_1}^{\beta_2} W_\sigma(u)Q_n(u)du + \sum_{i=1}^{2} W_{\sigma,\beta_i}Q_n(\beta_i), \tag{2.33}$$

which suggests estimating μ and σ by

$$\begin{bmatrix}\hat{\mu}\\ \hat{\sigma}\end{bmatrix} = \hat{I}^{-1}(\beta_1,\beta_2)\begin{bmatrix}\langle f_0Q_0, f_0Q_0\cdot Q_n\rangle_{\hat{K}}\\ \langle f_0Q_0\cdot Q_0, f_0Q_0\cdot Q_n\rangle_{\hat{K}}\end{bmatrix}. \tag{2.34}$$

Our principal result, given in Theorem 1, is that asymptotically $(\hat{\mu},\hat{\sigma})^t$ has the same distribution as $(\mu^*,\sigma^*)^t$.

<u>Theorem 1</u>. Assume that

(i) F is a strictly increasing continuously differentiable function, with $F(0) = 0$, H is a continuously differentiable function which satisfies $HQ(\beta_2) < 1$, and

(ii) both f_0Q_0 and $f_0Q_0\cdot Q_0$ are twice continuously differentiable functions that are elements of the RKHS generated by the Brownian bridge covariance kernel $\min(s,t)-st$.

Under the regularity conditions (i) and (ii)

$$\sqrt{n}\left\{\begin{bmatrix}\hat{\mu}\\ \hat{\sigma}\end{bmatrix}-\begin{bmatrix}\mu\\ \sigma\end{bmatrix}\right\} \to_L N(\underline{0},\sigma^2 I^{-1}(\beta_1,\beta_2)),$$

where $\to_L$ denotes covergence in law and $N(\underline{0},\sigma^2 I^{-1}(\beta_1,\beta_2))$ is a normal distribution with mean $\underline{0}$ and variance-covariance matrix $\sigma^2 I^{-1}(\beta_1,\beta_2)$.

Remark 1. Throughout this section we have tacitly treated the censoring distribution, H,as if it were known. This will frequently not be the case in practice. Such difficulties can be alleviated by substituting strongly consistent estimators for H and h in (2.26) - (2.31). It will be apparent from the proof in the next section that Theorem 1 still holds for estimators obtained in this manner. One simple approach to the estimation of H and h can be developed as follows. Estimators for expressions involving HQ can be based on $1-F_n^*Q_n(u)$ (F_n^* denoting the empirical d.f. for $Z_1,\ldots,Z_n$) which estimates $(1-u)(1-HQ(u)) = 1-F^*Q(u)$. Then through the use of differences, as in Sacks (1975) and Weiss and Wolfowitz (1970), one can construct an estimator for $(HQ)'(u)= \sigma hQ(u)/f_0Q_0(u)$ (Note that this allows us to dispense with $\hat{\sigma}$ in (2.24)). It would be of interest to see if one can extend this type of approach to include the estimation of f_0Q_0 and $f_0Q_0\cdot Q_0$ thereby obtaining censored sample analogs of the totally nonparametric estimators given in Weiss and Wolfowitz (1970). These latter types of estimators will be considered in more detail in future work.

Remark 2. We note that the conditions on f_0Q_0 and $f_0Q_0\cdot Q_0$ in Theorem 1 are implicit in the estimators developed by Parzen (1979a) for the uncensored case and are satisfied for most distributions. The assumption that $HQ(\beta_2) < 1$ appears to be standard in the work on estimation from randomly censored data (c.f. Sander (1975a) and Reid (1981)).

Remark 3. The difficulties with extending our results to $\beta_1=0$, $\beta_2=1$ stem from the fact that the distribution theory for Q_n is

currently available only over $[0,\beta_2]$. For Theorem 1 to apply when $\beta_1 = 0$ we would need to assume, as in Sander (1975a), that $f_0Q_0(0) \neq 0$ which is seldom satisfied. By requiring $\beta_1 > 0$ this assumption is averted and estimators that have wider applicability are obtained. In particular this formulation is well suited to dealing with estimation from truncated samples. Such an approach was also taken by Weiss (1964) and Weiss and Wolfowitz (1970).

To conclude this section we consider several examples that further illustrate the usefulness of model (2.5) and establish the connection between the estimators (2.34) and others already available. We begin by considering the form of $\hat{\mu}$ and $\hat{\sigma}$ in the case of no censorship.

Example 2. Consider the case of no censoring, i.e., assume H=0 or, more generally, that H places mass one at some point to the right of the support of F (or at $+\infty$). In this event $\langle m,g\rangle_K = \langle m,g\rangle_{\hat{K}}$ so that $I(\beta_1,\beta_2) = \hat{I}(\beta_1,\beta_2)$ and the weights (2.26) - (2.31) reduce to

$$W_\mu(u) = -(f_0Q_0)''(u)f_0Q_0(u) \tag{2.35}$$

$$W_{\mu,\beta_1} = [\frac{1}{\beta_1}f_0Q_0(\beta_1)-(f_0Q_0)'(\beta_1)]f_0Q_0(\beta_1) \tag{2.36}$$

$$W_{\mu,\beta_2} = [\frac{1}{1-\beta_2}f_0Q_0(\beta_2) + (f_0Q_0)'(\beta_2)]f_0Q_0(\beta_2) \tag{2.37}$$

$$W_\sigma(u) = -(f_0Q_0\cdot Q_0)''(u)f_0Q_0(u) \tag{2.38}$$

$$W_{\sigma,\beta_1} = [\frac{1}{\beta_1}f_0Q_0(\beta_1)Q_0(\beta_1)-(f_0Q_0\cdot Q_0)'(\beta_1)]f_0Q_0(\beta_1) \tag{2.39}$$

and

$$W_{\sigma,\beta_2} = [\frac{1}{1-\beta_2}f_0Q_0(\beta_2)Q_0(\beta_2)+(f_0Q_0\cdot Q_0)'(\beta_2)]f_0Q_0(\beta_2). \tag{2.40}$$

These are precisely the weights for the estimators proposed by Parzen (1979a, 1979b) and consequently $(\hat{\mu},\hat{\sigma})^t$ agrees with his estimators in the case of no censoring. Of course, this is no surprise since the covariance kernel (2.3) becomes the Brownian bridge covariance kernel when H = 0 and, hence, model (2.5) reduces to Parzen's model when no censoring is present (in fact,

this work was suggested by the survival data problem in Section 12 of Parzen (1979a)). The results of this section can also be viewed as a censored sample extension of work by Weiss (1964) due to the similarity between his estimators and those provided by (2.34) when H = 0.

Example 3. Eubank and LaRiccia (1981) consider a minimum distance approach to parameter estimation from censored data. For the estimation of μ when σ is known and without loss of generality taken to be one, their estimator is obtained by minimizing, with respect to μ, the distance

$$\int_{\beta_1}^{\beta_2} [Q_n(u)-\mu-Q_0(u)]^2 \phi(u)du + \sum_{i=1}^{2} \phi_{\beta_i} [Q_n(\beta_i)-\mu-Q_0(\beta_i)]^2 \ , \quad (2.41)$$

where ϕ, ϕ_{β_1} and ϕ_{β_2} are user selected weights. They give optimal weights for the estimation of a single parameter that, in this case, are

$$\phi(u) = -\{(f_0Q_0)''(u)[1-HQ(u)] - A(u;f_0Q_0)\}f_0Q_0(u) \quad (2.42)$$

$$\phi_{\beta_1} = \{(f_0Q_0)(\beta_1)B(\beta_1) - (f_0Q_0)'(\beta_1)[1-HQ(\beta_1)]\}f_0Q_0(\beta_1) \quad (2.43)$$

and

$$\phi_{\beta_2} = \{\frac{1}{1-\beta_2} f_0Q_0(\beta_2) + (f_0Q_0)'(\beta_2)\}[1-HQ(\beta_2)]f_0Q_0(\beta_2) \quad (2.44)$$

where, recall, $Q(u) = \mu + Q_0(u)$ and $\sigma = 1$ in $A(u;f_0Q_0)$.

As (2.42) - (2.44) depend upon the unknown parameter, μ, an iterative procedure is required to obtain the estimator that minimizes (2.41). However, if one replaces Q by Q_n in (2.42) - (2.44) and formally differentiates (2.41) the resulting estimator is precisely $\hat{\mu}$. Thus one consequence of this paper is that, for location parameter estimation, the asymptotic properties of the minimum distance procedure involving (2.41) still hold when the optimal weights are estimated.

It is of interest to note that for more general parameterizations than those of the location-scale variety a model such as (2.5) will still hold that involves the parameters in a nonlinear

fashion. Thus minimum distance procedures such as those in Eubank and LaRiccia (1981) may be viewed, intuitively, as providing parameter estimates via nonlinear regression.

Finally we note that for the estimation of either μ or σ when $H = 0$ the minimum distance procedures of Eubank and LaRiccia (1981) and LaRiccia (1982) when using optimal weights, and the estimators given by Parzen (1979a, b) will all three coincide.

Example 4. Frequently for computational as well as other reasons one may wish to use only a subsample consisting of k sample quantiles for the estimation of μ and σ. A popular estimator of this type for the uncensored case has been the asymptotically best linear unbiased estimator (ABLUE) developed by Ogawa (1951). We now show how an analogous estimator can be obtained for censored samples.

Let $0 < \beta_1 \leq u_1 < u_2 < \ldots < u_k \leq \beta_2 < 1$ denote the percentile points of the quantiles to be utilized in the estimator. As in the uncensored case $U = \{u_1,\ldots,u_k\}$ will be termed a spacing. The problem of estimating μ and σ from the quantile subset corresponding to U can be viewed as estimation using a sample obtained from model (2.5) by sampling at the elements $u_1,\ldots,u_k$. Therefore, let K_U denote the $k\times k$ matrix with ij^{th} element $K(u_i,u_j)$ and define the $k\times 2$ matrix $M_U = \{g_j(u_i)\}$ where $g_1 = f_0Q_0$ and $g_2 = f_0Q_0\cdot Q_0$. Then, using Y to denote the $k\times 1$ observation vector $(f_0Q_0(u_1)Q_n(u_1),\ldots,f_0Q_0(u_k)Q_n(u_k))^t$, an (asymptotically) best linear unbiased estimator is

$$\begin{bmatrix} \mu^*(U) \\ \sigma^*(U) \end{bmatrix} = (M_U^tK_U^{-1}M_U)^{-1}M_U^tK_U^{-1}\underline{Y} \tag{2.45}$$

(that this estimator is well defined follows from Sacks and Ylvisaker (1966)).

For the Koziol-Green model the elements of K_U are

$$K(u_i,u_j) = (1-u_j)[(1-u_i)^\alpha - (1-u_i)]/(1+\alpha)$$

but will, in general, depend on the unknown parameters. However, we can estimate $K(u_i,u_j)$ as in (2.20) and it will follow from work in the next section that the estimator $\hat{K}_U$ obtained in this manner converges in probability to K_U. Straightforward arguments then show that an estimator with the same asymptotic distribution as $(\mu^*(U), \sigma^*(U))^t$ is

$$\begin{bmatrix} \hat{\mu}(U) \\ \hat{\sigma}(U) \end{bmatrix} = (M_U^t \hat{K}_U^{-1} M_U)^{-1} M_U^t \hat{K}_U^{-1} \underline{Y}. \tag{2.46}$$

Both of (2.45) and (2.46) reduce to the ABLUE when $H = 0$.

Considerable attention has been focused on the problem of optimal quantile (spacing) selection for the ABLUE (see Eubank (1981) for references). Similar problems can be considered here. For instance under the Koziol-Green model (2.45) depends only on the possibly unknown parameter α. One might therefore consider the selection of U to maximize $\det(M_U^t K_U^{-1} M_U)$ over various values for α.

3. PROOF OF THEOREM

Let

$$\underline{Z} = \begin{bmatrix} \langle f_0 Q_0, f_0 Q_0 \cdot Q_n \rangle_K \\ \langle f_0 Q_0 \cdot Q_0, f_0 Q_0 \cdot Q_n \rangle_K \end{bmatrix} \tag{3.1}$$

and define

$$\hat{\underline{Z}} = \begin{bmatrix} \langle f_0 Q_0, f_0 Q_0 \cdot Q_n \rangle_{\hat{K}} \\ \langle f_0 Q_0 \cdot Q_0, f_0 Q_0 \cdot Q_n \rangle_{\hat{K}} \end{bmatrix} . \tag{3.2}$$

To prove Theorem 1 we show, sequentially, that

A. $\hat{I}(\beta_1,\beta_2) \to_P I(\beta_1,\beta_2)$

B. $\sqrt{n}\{\underline{Z} - I(\beta_1,\beta_2)\begin{bmatrix}\mu \\ \sigma\end{bmatrix}\} \to_L N(\underline{0},\sigma^2 I(\beta_1,\beta_2))$

and that

C. $\sqrt{n}\{\hat{I}^{-1}(\beta_1,\beta_2)\hat{\underline{Z}} - I^{-1}(\beta_1,\beta_2)\underline{Z}\} \to_{\underline{P}} \underline{0}$.

As A and C together imply that

$$\sqrt{n}\left\{\begin{bmatrix}\hat{\mu} \\ \hat{\sigma}\end{bmatrix} - \begin{bmatrix}\mu \\ \sigma\end{bmatrix}\right\} = o_p(1) + I^{-1}(\beta_1,\beta_2)\sqrt{n}\left\{\underline{Z} - I(\beta_1,\beta_2)\begin{bmatrix}\mu \\ \sigma\end{bmatrix}\right\} , \tag{3.3}$$

where $o_p(1) \to_P 0$ as $n\to\infty$, the theorem then follows as a consequence of B.

The fundamental results that will be needed regarding the closeness of Q_n and Q are furnished by the following lemma.

<u>Lemma 1.</u> Let condition i) of Theorem 1 be satisfied and define

$$\rho(Q_n,Q) = \sup_{\beta_1\le u\le\beta_2} |Q_n(u) - Q(u)|.$$

Then,

i) $\sqrt{n}\,\rho(Q_n,Q) = O_p(1)$,
where $O_p(1)$ denotes bounded in probability as $n\to\infty$ and

ii) if g is a continuous function on $[Q(\beta_1), Q(\beta_2)]$, $\rho(g(Q_n), g(Q)) \to_P 0$.

<u>Proof</u>: It is shown in Sander (1975a) that $\sup_{u\varepsilon[\beta_1,\beta_2]} |X(u)|$ is a random variable. Since the supremum is a continuous function standard arguments show i). Also note, from i), that $\rho(Q_n,Q) = \frac{1}{\sqrt{n}}(\sqrt{n}\rho(Q_n,Q)) \to_P 0$.

To prove ii) first let $\Lambda_n = \{\omega\colon Q_n(u)$ is well defined on $[\beta_1,\beta_2]$ and $Q_n(\beta_1) \ge Q(\beta_1)$, $Q_n(\beta_2) \le Q(\beta_2)\}$. By an argument similar to that used in proving Lemma 3 of Sander (1975a) which employs the consistency of F_n it follows that $P(\Lambda_n) \to 1$. Consequently, it suffices to prove ii) on Λ_n. However, on Λ_n the result is an immediate consequence of the fact that $\rho(Q_n,Q) \to_P 0$ and the uniform continuity of g on $[Q(\beta_1), Q(\beta_2)]$.

To verify A let $g_1 = f_0Q_0$, $g_2 = f_0Q_0\cdot Q_0$ and observe that from (2.6) and (2.19)

$$\langle g_i,g_j\rangle_{\hat{K}} - \langle g_i,g_j\rangle_K = \int_{\beta_1}^{\beta_2} \left\{\frac{g_i(u)}{1-u}\right\}' \left\{\frac{g_j(u)}{1-u}\right\}' (1-u)^2[HQ(u)-HQ_n(u)]du + \frac{g_i(\beta_1)g_j(\beta_1)}{1-\beta_1}\left[\frac{1}{\hat{G}(\beta_1)} - \frac{1}{G(\beta_1)}\right]. \tag{3.4}$$

The integral in (3.4) is seen to be $o_p(1)$ using part ii) of the lemma, the continuity of H and condition ii) of Theorem 1, since $\int_{\beta_1}^{\beta_2}\left\{\frac{g_i(u)}{1-u}\right\}'\left\{\frac{g_j(u)}{1-u}\right\}' du$ is essentially the inner product for the RKHS generated by min(s,t)-st. It then remains to show that $\left|\frac{1}{\hat{G}(\beta_1)} - \frac{1}{G(\beta_1)}\right| \to_P 0$. This can be seen by using Lemma 1 of Sander (1975a) to show that $\hat{G}(\beta_1)$ has the same asymptotic distribution as

$$G_1^*(\beta_1) = (1-\beta_1)\int_0^{Q_n(\beta_1)} \frac{1}{[1-F_n(x)]^2[1-H(x)]}\, dF_n(x) \ . \qquad (3.5)$$

The convergence of $\hat{G}_1(\beta_1)$ to $G(\beta_1)$ then follows using arguments similar to those in Breslow and Crowley (1974). Thus, $1/\hat{G}(\beta_1)$ is bounded in probability and

$$\left|\frac{1}{\hat{G}(\beta_1)} - \frac{1}{G(\beta_1)}\right| = O_p(1)\left|G(\beta_1) - \hat{G}(\beta_1)\right| \to_P 0.$$

This proves A.

To obtain B we use the weak convergence properties of $\sqrt{n}f_0Q_0(u)\{Q_n(u)-Q(u)\}$ on $[\beta_1,\beta_2]$ and the continuous mapping theorem (c.f. Billingsley (1968)) to conclude that, asymptotically, $\underline{Z}$ has a normal distribution with mean

$$\underline{m} = \begin{bmatrix} \langle f_0Q_0, f_0Q_0\cdot Q\rangle_K \\ \langle f_0Q_0\cdot Q_0, f_0Q_0\cdot Q\rangle_K \end{bmatrix} = I(\beta_1,\beta_2)\begin{bmatrix}\mu \\ \sigma\end{bmatrix} \qquad (3.6)$$

as a result of (2.1). Now, from (2.8) and the reproducing property for K, namely $g(u) = \langle K(\cdot,u), g\rangle_K$, f_0Q_0 is found to admit the representation

$$f_0Q_0(u) = \int_{\beta_1}^{\beta_2}\phi(s)K(s,u)du + \sum_{i=1}^{2}\phi_{\beta_k}K(\beta_k,u) \qquad (3.7)$$

where

$$\phi(s) = -(f_0Q_0)''(u)[1-HQ(u)] - A(u;f_0Q_0) \qquad (3.8)$$

$$\phi_{\beta_1} = f_0Q_0(\beta_1)B(\beta_1) - (f_0Q_0)'(\beta_1)[1-HQ(\beta_1)] \tag{3.9}$$

and

$$\phi_{\beta_2} = \{\frac{1}{1-\beta_2} f_0Q_0(\beta_2) + (f_0Q_0)'(\beta_2)\}[1-HQ(\beta_2)]. \tag{3.10}$$

It then follows from (3.7) and another application of the reproducing property that

$$||f_0Q_0||_K^2 = \int_{\beta_1}^{\beta_2}\int_{\beta_1}^{\beta_2}\phi(u)\phi(v)K(u,v)dudv + \sum_{k=1}^{2}\phi_{\beta_k}\int_{\beta_1}^{\beta_2}\phi(u)K(u,\beta_k)du$$

$$+ \sum_{k=1}^{2}\sum_{\ell=1}^{2}\phi_{\beta_k}\phi_{\beta_\ell}K(\beta_k,\beta_\ell) \quad . \tag{3.11}$$

Equation (3.11) apart from the constant multiple σ^2, is precisely the variance of $<f_0Q_0, f_0Q_0\cdot Q_n>_K$ obtained by application of the continuous mapping theorem. Analogous results are seen to hold for the remainder of the elements of $I(\beta_1,\beta_2)$ through identical arguments. Apparently B can also be obtained using techniques such as those in Example 3 of Reid (1981) although more stringent conditions are required on the weights.

Finally to prove C note that, just as in (3.6),

$$\hat{\underline{m}} = \begin{bmatrix} <f_0Q_0, f_0Q_0\cdot Q>_{\hat{K}} \\ <f_0Q_0\cdot Q_0, f_0Q_0\cdot Q>_{\hat{K}} \end{bmatrix} = \hat{I}(\beta_1,\beta_2)\begin{bmatrix}\mu \\ \sigma\end{bmatrix} \quad . \tag{3.12}$$

Thus

$$\sqrt{n}\left\{\begin{bmatrix}\hat{\mu} \\ \hat{\sigma}\end{bmatrix} - \begin{bmatrix}\mu^* \\ \sigma^*\end{bmatrix}\right\} = \hat{I}^{-1}(\beta_1,\beta_2)\ \sqrt{n}[\hat{\underline{d}}-\underline{d}] - [I^{-1}(\beta_1,\beta_2)-\hat{I}^{-1}(\beta_1,\beta_2)]\sqrt{n}\underline{d}$$

where $\hat{\underline{d}} = \hat{\underline{Z}} - \hat{\underline{m}}$ and $\underline{d} = \underline{Z} - \underline{m}$. From A and B we conclude that $[I^{-1}(\beta_1,\beta_2)-\hat{I}^{-1}(\beta_1,\beta_2)]\sqrt{n}\underline{d} \to_P \underline{0}$ so it remains only to show that $\sqrt{n}[\hat{\underline{d}}-\underline{d}] \to_P \underline{0}$. This can now be verified directly through tedious but straightford arguments that require the use of the continuity of h and H, the fact that $\hat{\sigma} \to_P \sigma$, arguments similar to those used in the proof of A and repeated applications of parts i) and ii) of the lemma. This completes the proof of C and, hence, the theorem.

ACKNOWLEDGEMENT

The research of the first author was partially supported by the Office of Naval Research Contract N00014-82-K-0207. Portions of the research of the second author were accomplished while at the University of Nebraska-Lincoln.

BIBLIOGRAPHY

Aronszajn, N. (1950). Theory of reproducing kernels. Trans. Amer. Math. Soc. 68, 337-404.

Billingsley, P. (1968). Weak Convergence of Probability Measures. Wiley, New York.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimator under random censorship. Ann. Statist. 2, 437-453.

Csörgő, M. and Révész, P. (1978). Strong approximation of the quantile process. Ann. Statist. 6, 882-894.

Csörgő, S. and Horváth, L. (1981). On the Koziol-Green model for random censorship. Biometrika 68, 391-401.

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. (1981). Weighted L^2 quantile distance estimators for randomly censored data. Tech. Rep. No. 149, Dept. of Statist., Southern Methodist University, Dallas, Texas.

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. Biometrics 35, 393-401.

Kalbfleisch, J. D. and Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-581.

Koziol, J. and Green, S. (1976). A Cramér-von Mises statistic for randomly censored data. Biometrika 6, 465-474.

LaRiccia, V. N. (1982). Asymptotic properties of weighted L^2 quantile distance estimators. Ann. Statist. 10, 621-624.

Ogawa, J. (1951). Contributions to the theory of systematic statistics I. Osaka Math. J. 3, 175-223.

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

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

Parzen, E. (1979a). Nonparametric statistical data modeling. J. Amer. Statist. Assoc. 74, 105-120.

Parzen, E. (1979b). A density-quantile function perspective on robust estimation. In Robustness in Statistics, edited by R. Lanner and G. Wilkinson. Academic Press, New York, 237-258.

Reid, N. (1981). Influence functions for censored data. Ann. Statist. 9, 78-92.

Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. Ann. Statist. 2, 285-298.

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

Sander, J. (1975a). The weak convergence of quantiles of the product-limit estimator. Tech. Rep. No. 5, Dept. of Statist., Stanford Univ.

Sander, J. (1975b). Asymptotic normality of linear combinations of functions of order statistics with censored data. Tech. Rep. No. 8, Dept. of Statist., Stanford Univ.

Sander, J. (1975c). A comparison of three estimators derived from the Kaplan-Meier estimator. Tech. Rep. No. 11, Dept. of Statist., Stanford Univ.

Susarla, V. and Van Ryzin, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. Ann. Statist. 8, 1001-1016.

Weiss, L. (1964). On estimating location and scale parameters from truncated samples. Naval Res. Logist. Quart. 11, 125-134.

Weiss, L. and Wolfowitz, J. (1970). Asymptotically efficient nonparametric estimators of location and scale parameters. Z. Wahrscheinlichkietstheorie verw. Geb. 16, 134-150.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 166

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): LOCATION AND SCALE PARAMETER ESTIMATION FROM RANDOMLY CENSORED DATA

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER: 166

7. AUTHOR(s): R. L. Eubank & V. N. LaRiccia

8. CONTRACT OR GRANT NUMBER(s): N00014-82-k-0207

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Southern Methodist University
Dallas, TX 75275
and
University of Delaware, Newark, Delaware 19711

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: August, 1982

13. NUMBER OF PAGES: 21

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 purposes 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)

Quantile funtion; estimation; location and scale parameters; random censoring; reproducing kernel Hilbert space.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of location and scale parameter estimation from randomly censored data is analyzed through use of a regression model for the Kaplan-Meier quantile process. Continuous time regression techniques are employed to construct estimators that are both asymptotically normal and efficient. Estimators with a particularly simple form are obtained for the Koziol-Green model for random censorship. In the event of no censoring the regression model, and resulting estimators, reduce to those proposed by Parzen (1979 a, b).

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

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)