REGRESSION QUANTILES
IN
NONPARAMETRIC REGRESSION

by

PRANAB KUMAR SEN

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2116

August 1993

REGRESSION QUANTILES IN NONPARAMETRIC REGRESSION

PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

In a nonparametric setup involving stochastic regressors, regression quantiles relate to the so called conditional quantile functions. Various asymptotic properties of such conditional quantile processes are studied with due emphasis on the underlying design aspects.

KEYWORDS. Conditional quantile function; ε -affine design; kernel method; nearest neighborhood; semi-parametric model; sequential design; smoothing.

1. INTRODUCTION

In a multivariate setup, suppose that $(Y_i, \underset{\sim}{Z}_i)$, $i=1,\ldots,n$ are independent and identically distributed (i.i.d.) random vectors (r.v.) with a $(q+1)$-dimensional distribution function (d.f.) $\Pi(.)$. In a conventional parametric framework, it is usually assumed that given $\underset{\sim}{Z}_i = \underset{\sim}{z}$ ($\in R^q$, for some $q \geq 1$), Y_i follows a linear model with mean $\theta + \underset{\sim}{\beta}'\underset{\sim}{z}$ and a finite and positive variance σ^2 , which does not depend on the specific realization $(\underset{\sim}{z})$. In this setup, θ is the intercept parameter, $\underset{\sim}{\beta}$ is the vector of regression parameters and σ^2 is the error variance. This linearity of the model and the homoscedasticity constitute the two basic assumptions, and in actual practice, there may be situations where none of these may be strictly true. Any departure from these model-assumptions , even to a minor extent, may cause considerable damage to the statistical analysis based on such models, i.e., the validity and efficacy of such statistical conclusions may be at a stake. The assumption of normality of the error component is crucial for claiming small (or finite) sample optimality properties of the estimators of these parameters and/or for suitable tests of hypotheses on them. In large samples, this normality assumption can be dispensed with to a certain extent and standard asymptotic theory can be incorporated in the derivation of suitable asymptotic optimality properties of the classical least squares estimators (LSE) and/or the normal theory maximum likelihood estimators (MLE) as well as the associated tests. However, such procedures are generally not robust against plausible departures from the linearity of the model and/or the homoscedasticity assumptions made before. For this reason, alternative robust estimators (viz., R-, L- and M-estimators of location and regression) have been proposed by a host of workers. These estimators relate to the so called semi-parametric model :

AMS Subject Classification: 62G05, 62G07, 62G10, 62G20.

Presented at the Third Eugene Lukacs Symposium at Bowling Green State University, Bowling Green, Ohio, March 25-27, 1993.

$$G(y \mid \underset{\sim}{z}) = P\{ Y \leq y \mid \underset{\sim}{Z}=\underset{\sim}{z} \} = G(y - \theta - \underset{\sim}{\beta}'\underset{\sim}{z}),\ y \in R,\ \underset{\sim}{z} \in R^q, \tag{1.1}$$

where G is an arbitrary (continuous) d.f. and it does not depend on $\underset{\sim}{z}$. In this formulation, the parametric part is retained through the linearity component and also the homoscedasticity part is retained through the independence of G(x) on the conditioning variate $\underset{\sim}{z}$, while the nonparametric flavor is imported through the arbitrariness of the d.f. G. Basically, such procedures are robust against possible departures from the normality of (either locally or globally), but are still motivated by the two basic assumptions in linear models, mentioned earlier.

There has been some attempts to accommodating possible heteroscedasticity in (1.1) through smoothing techniques based on the nearest neighbor (NN-)methods and the popular kernel methods. In this setup, it is often assumed that

$$E[\ Y \mid \underset{\sim}{Z} = \underset{\sim}{z}\ ] = m(\underset{\sim}{z}) \quad \text{and} \quad Var(Y \mid \underset{\sim}{Z}=\underset{\sim}{z}) = \sigma^2(\underset{\sim}{z}),\ \underset{\sim}{z} \in R^q\ , \tag{1.2}$$

where m(.) and $\sigma^2(.)$ may depend on $\underset{\sim}{z}$ quite arbitrarily, but satisfy some local smoothness properties which enable the NN- or kernel methods to work out the statistical solutions neatly. Bhattacharya (1963) and Stone (1977) are the most notable references in this direction. As in (1.1), we may denote the conditional d.f. of Y, given $\underset{\sim}{Z} = \underset{\sim}{z}$, by $G(y \mid \underset{\sim}{z})$, and instead of the conditional expectation in (1.2), we may also formulate a functional

$$\theta(\underset{\sim}{z}) = \inf\{\ y : G(y \mid \underset{\sim}{z}) \geq\ p\ \}\ , \text{ for some } p \in (0,1), \tag{1.3}$$

so that $\theta(\underset{\sim}{z})$ is the conditional p-quantile of Y given $\underset{\sim}{Z} = \underset{\sim}{z}$. Such a conditional quantile function demands the least restrictive assumptions on the basic model and allows possible heteroscedasticity in a very natural manner. We may refer to Bhattacharya and Gangopadhyay (1990) for a nice account of the general asymptotic theory of sample estimates of the $\theta(\underset{\sim}{z})$. Gangopadhyay and Sen (1992) considered a formulation relating to a general statistical functional $\tau(G(.|\underset{\sim}{z}))$ of the conditional d.f. $G(.|\ \underset{\sim}{z})$ which includes (1.2) and (1.3) as special cases. In this framework one has a complete nonparametric formulation of the regression problem, and under suitable differentiability conditions on $\tau(.)$, the asymptotic properties of the sample estimates have been studied by them. Contiguity of probability measures, as extended to such conditional laws, plays a vital role in this context. Typically, $\theta(\underset{\sim}{z})$ or $\tau(G(.|\underset{\sim}{z}))$, being a function of $\underset{\sim}{z}$, need to be estimated for a range of values of $\underset{\sim}{z}$, so that in this generality, we have a conditional quantile process , which have been studied in detail by Gangopadhyay and Sen (1993). It is quite appropriate to refer to these as regression quantiles (functions) in a complete nonparametric framework. The primary object of the current investigation is to study some general properties of such regression quantiles with due emphasis on their relevance to the other types of regression quantiles mostly developed for nonstochastic regressors.

2. REGRESSION QUANTILES FOR NONSTOCHASTIC REGRESSORS

Consider the usual linear regression model :

$$\underset{\sim}{Y} = (Y_1,\ldots,Y_n)' = \underset{\sim\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} \; ; \; \underset{\sim}{e} = (e_1,\ldots, e_n)' \; , \tag{2.1}$$

where $\underset{\sim}{X}$ is an $n\times p$ matrix of known regression constants, $\underset{\sim}{\beta} = (\beta_1,\ldots, \beta_p)'$ is a vector of unknown regression parameters, and the e_i are i.i.d.r.v.'s having a continuous d.f. F , defined on R. Koenker and Bassett (1978) introduced the idea of regression quantiles for the model (2.1). For a given $\gamma : 0 < \gamma < 1$, $\hat{\underset{\sim}{\beta}}^*_n(\gamma)$, the γ-**regression quantile** of $\underset{\sim}{\beta}$, can be expressed as a solution of the minimization of

$$\Sigma_{i=1}^n \; \rho_\gamma(Y_i - \underset{\sim}{x}_i'\underset{\sim}{b}) \; , \; \underset{\sim}{b} \in R^p \; , \tag{2.2}$$

where $\underset{\sim}{X}' = (\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n)$, and

$$\rho_\gamma(x) = |x|\{ \quad \gamma I(x > 0) + (1- \gamma)I(x < 0)\}, \; x \in R. \tag{2.5}$$

Koenker and Bassett (1978) also characterized $\hat{\underset{\sim}{\beta}}^*_n(\gamma)$ as the component of $\underset{\sim}{\beta}$ in the optimal solution $(\underset{\sim}{\beta}, \underset{\sim}{r}^+ ,\underset{\sim}{r}^-)$ of the linear programme

$$\begin{aligned} &\gamma \underset{\sim n}{1}'\underset{\sim}{r}^+ + (1 - \gamma)\underset{\sim n}{1}'\underset{\sim}{r}^- = \text{minimum}; \; \underset{\sim n}{1} = (1,\ldots,1)' \; ; \\ &\underset{\sim\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{r}^+ - \underset{\sim}{r}^- = \underset{\sim}{Y} \; ; \; (\underset{\sim}{\beta}, \underset{\sim}{r}^+, \underset{\sim}{r}^-) \in R^p \times R^{n+} \times R^{n+} \; . \end{aligned} \tag{2.6}$$

Such regression quantiles are extensions of the conventional sample quantiles in the one sample case to the linear model in (2.1), and they provide a basis for L-estimation in linear models. We may refer to a recent paper of Gutenbrunner and Jurečková (1992) where a detailed treatment of regression quantiles is available, and other pertinent references are also cited there. The dual programme to (2.6) can be written as

$$\underset{\sim}{Y}'\hat{\underset{\sim n}{a}}(\gamma) = \max \; , \; \underset{\sim}{X}'\hat{\underset{\sim n}{a}}(\gamma) = (1- \gamma)\underset{\sim}{X}'\underset{\sim n}{1}; \; \hat{\underset{\sim n}{a}}(\gamma) \in [0,1]^n, \tag{2.7}$$

where $\gamma \in (0,1)$. Because of their alience to the Hájek-Šidák (1967,Ch.V) rank scores (in the one sample model), Gutenbrunner and Jurečková (1992) termed the $\hat{\underset{\sim n}{a}}(\gamma)$ as the (vector of) **regression rank scores**. One important property of these regression rank scores is their **regression-invariance**, i.e.,

$$\hat{\underset{\sim n}{a}}(\gamma; \underset{\sim}{Y} + \underset{\sim\sim}{X}\underset{\sim}{b}) = \hat{\underset{\sim n}{a}}(\gamma; \underset{\sim}{Y}), \text{ for every } \underset{\sim}{b} \in R^p \; . \tag{2.8}$$

This regression-invariance property makes it very convenient (and comparable to the LSE) when the linear model in (2.1) holds. For general non-linear models, Jurečková (1993) has worked out the regression quantile estimators and studied their asymptotic properties. However, in this context, generally the aforementioned invariance property may not hold. In either case of linear or non-linear models, the basic fact is that the regression functions are dictated by a finite set of parameters, and hence, a finite-dimensional analysis suffices. This basic feature is generally

untenable in a completely nonparametric regression model [such as in (1.2) or (1.3)] where the regression functions are generally not parameterizable in terms of a fixed set of parameters in a specified functional form. This is the main reason why some local smoothing techniques are usually employed to reduce the dimensionality, albeit at a cost of a somewhat slower rate of convergence and often with a comparable rate of the associated bias function. Keeping in mind this scenario of nonparametric regression models, we shall confine ourselves to such alternative procedures with due emphasis on the regression quantiles in a purely nonparametric setup.

3. ESTIMATION OF CONDITIONAL QUANTILE FUNCTIONS

For the set of stochastic vectors $(Y_i, \underset{\sim}{Z}_i)$, i=1,...,n, introduced in Section 1, a conditional quantile function $\theta(\underset{\sim}{z})$, $\underset{\sim}{z} \in R^q$, has already been defined in (1.3). In this formulation, we do not need to impose parametric structures on $\theta(\underset{\sim}{z})$, although we may like to introduce local smoothness conditions under which more effective and efficient statistical analysis can be performed. In this context, both the NN- and kernel methods work out and their asymptotic relative performance are very much comparable. As in Gangopadhyay and Sen (1992,1993), we shall find it more convenient to describe the procedure based on the NN-method, and with little modifications it amends to the kernel method as well; these will be briefly pointed out too.

The NN-methodology relates to point-wise estimation of $\theta(\underset{\sim}{z})$, so that a conditional quantile function has to be constructed from the accumulation of all such pivotal points belonging to a region which pertains to the basic domain of support of the covariates $\underset{\sim}{Z}_i$. A neighborhood is defined with respect to a suitable norm or metric, and in the general case of q-vectors , for some $q \geq 1$, some care need to be taken in adopting such a norm.For q = 1, a neighborhood may be an one-sided interval with the pivot as an extreme point or usually a two-sided interval with the pivot as the centre. In general, for every $\underset{\sim}{z}_0 \in R^q$ and $\underset{\sim}{z} \in R^q$, we conceive of a metric

$$\rho(\underset{\sim}{z},\underset{\sim}{z}_0) : R^q \times R^q \rightarrow R^+ , \tag{3.1}$$

satisfying the usual regularity properties of a norm. It may be useful to keep in mind the usual Euclidean norm ($= ||\underset{\sim}{z} - \underset{\sim}{z}_0|| = (\{(\underset{\sim}{z}-\underset{\sim}{z}_0)'(\underset{\sim}{z}-\underset{\sim}{z}_0)\})^{1/2}$), although we may like to have more general quadratic norms incorporating suitable weight matrices to reflect the relative importance of the various coordinates of $\underset{\sim}{z}$ as well as their statistical interdependence. With respect to a chosen $\underset{\sim}{z}_0$, consider the set of random elements :

$$\rho(\underset{\sim}{Z}_i,\underset{\sim}{z}_0) = D_i \text{ , say, for } i=1,\ldots,n. \tag{3.2}$$

The D_i are nonnegative i.i.d.r.v.'s with a d.f. depending on the joint d.f. of $\underset{\sim}{Z}$ as well as the choice of the metric $\rho(.)$. In a NN-method, corresponding to the sequence of sample size (n), we choose a sequence $\{k_n\}$, such that k_n is increasing in n and

$n^{-1}k_n$ converges to 0 as $n \to \infty$. We denote the order statistics corresponding to the D_i in (3.2) by $D_{n:1} \leq \ldots \leq D_{n:n}$ respectively. If the distribution of D is continuous, ties among the D_i (and hence, $D_{n:i}$) may be neglected with probability one. With this, we define (a stochastic) neighborhood of the pivot $\underset{\sim}{z}_0$ by

$$Nhd(\underset{\sim}{z}_0;\ \rho,\ k_n\ ) = \{\ \underset{\sim}{z} \in R^q : \rho(\underset{\sim}{z},\underset{\sim}{z}_0) \leq D_{n:k_n}\ \}\ . \tag{3.3}$$

[For example, with respect to the Euclidean norm, (3.3) relates to a sphere with center $\underset{\sim}{z}_0$ and radius $D_{n:k_n}$.] If D has a nondegenrate d.f., it follows under fairly general regularity conditions that $D_{n:k_n}$ converges to 0 almost surely (a.s.) as $n \to \infty$. In passing, we may remark that the $D_{n:i}$ as well as the neighborhood in (3.3) are highly dependent on the pivot $\underset{\sim}{z}_0$ and the metric $\rho(.)$.

Now corresponding to a chosen $\rho(.)$ and a given $\underset{\sim}{z}_0$, we define the anti-ranks S_i , i=1,...,n, by letting $D_{n:i} = D_{S_i}$, for i = 1,...,n. Thus, $(S_1,\ldots,S_n)$ is a random permutation of the numbers (1,...,n). Consider then the set of observations:

$$(Y_{S_i},\ \underset{\sim}{Z}_{S_i})\ ,\ i=1,\ldots,\ k_n\ . \tag{3.4}$$

Then the k_n-nearest neighbor (k_n-NN) empirical d.f. of Y with respect to z_0 and $\rho(.)$ is defined as

$$G_{nk_n}(y) = k_n^{-1}\ \Sigma_{i=1}^{k_n}\ I(Y_{S_i} \leq y\ ),\ y \in R\ , \tag{3.5}$$

where I(A) stands for the indicator function of the set A. Parallel to (1.3), we may then define the k_n-NN estimator of $\theta(\underset{\sim}{z}_0)$ by

$$\hat{\theta}_{nk_n}(\underset{\sim}{z}_0\ ) = \inf\{\ y : G_{nk_n}(y) \geq [pk_n]/k_n\ \}\ \ ;\ p \in (0,1). \tag{3.6}$$

There is a very natural interpretation of the estimator in (3.6). The Y_{S_i} are the induced order statistics or concomitants of order statistics in an extended sense, and the Bhattacharya (1974) lemma yields that the Y_{S_i} are conditionally (given the $\underset{\sim}{Z}_{S_i}$) independent r.v.'s, although they are not necessarily identically distributed. However, the very construction of the neighborhood in (3.3) makes the marginal (conditional) distributions of the Y_{S_i} in this set nearly identical,a fact that has been established through contiguity of (conditional) probability measures by Gangopadhyay and Sen (1992,1993). As such, if we consider the conditional law of Y given $\underset{\sim}{Z} = \underset{\sim}{z}_0$ and treat as if all the Y_{S_i} in this neighborhood have this distribution for which the conditional quantile is $\theta(\underset{\sim}{z}_0)$, then $\hat{\theta}_{nk_n}(\underset{\sim}{z}_0)$ is the natural sample counterpart of $\theta(\underset{\sim}{z}_0)$. For simplicity of presentation, Gangopadhyay and Sen (1992,1993) and most of the other workers in this area considered exclusively the case of real valued Z_i, and under appropriate regularity conditions on the marginal density of Z as well as the conditional density of Y given Z = z, various asymptotic properties of the estimator in (3.6) have been studied. Point-wise asymptotic normality and optimal choice of k_n in (3.3) constitute the main objective of such studies.

From the very basic definition of $\theta(z)$, even when z is real valued, it is quite clear that one has to consider the function in a general mold as

$$\theta(C) = \{ \theta(\underset{\sim}{z}) : \underset{\sim}{z} \in C \}, \tag{3.7}$$

where C is (usually a compact) subset of R^q which includes the essential support of the distribution of the concomitant variate $\underset{\sim}{Z}$. In this way, effective statistical conclusions can be drawn on the conditional quantiles for a range of concomitant variates. Even for the particular case of q = 1 and C,a compact interval on R, there are technical points in the construction of estimates of $\theta(C)$ in (3.7). The main difficulty is due to the fact that the estimator in (3.6) [or other rival ones] are are atuned to the specific point $\underset{\sim}{z}_0$ and the construction of the D_i in (3.2) is also linked to this pivot $\underset{\sim}{z}_0$. Thus, technically, one needs to consider all possible points $(\underset{\sim}{z}_0)$ in the set C, and repeat this procedure point-wise. This results into an infinite dimensional estimation problem and therefore the related asymptotics may require some extra care. For q = 1, Gangopadhyay and Sen (1993) attacked the problem using the same notion of contiguity (as employed in their earlier work). The weak convergence of such a conditional quantile function process is established under essentially the same regularity conditions as in the case of a single $\underset{\sim}{z}_0$. In the next section, we explore the situation with respect to the general model in (1.3). In the rest of this section, we motivate the relevance of design aspects in such studies. Recall that we started with a set of n i.i.d.r. vectors $(Y_i, \underset{\sim}{Z}_i)$, and then, for a given $\underset{\sim}{z}_0$, we considered the neighborhood in (3.3) based on which the estimator in (3.6) has been proposed and justified. However, since there is an infinite (i.e., uncountable) number of points $(\underset{\sim}{z}_0)$,even in a compact set $C \subset R^q$, a multi-dimensional time-parameter process,as can be obtained from (3.6) by allowing $\underset{\sim}{z}_0$ to vary over C , is essentially generated by a finite set of n random vectors. Hence, a finite dimensional statistical analysis should be capable of covering such an infinite dimensional problem, and in this context, suitable 'gridding' of the $\underset{\sim}{z}_0$ over the set C should provide us with the essential tools in performing such a finite dimensional statistical analysis. We refer to this aspect as the design of grid-points, and in a later section, we study the relevance of such designs in reducing the dimensionality of the problem. In this context too, one may need suitable smoothness conditions on the $\theta(\underset{\sim}{z})$, but these are very similar to the usual 'spline methodology' where only local smoothness patterns are imposed. For real valued Z_i, Gangopadhyay and Sen (1993) have partially explored this quantile process modeling, although very little has been stressed on the designing aspects. As such, we intend to present a general overview of the conditional quantile process surfaces in the q-variate case, and then to discuss the designing aspects in greater details for the real valued Z_i. We assume that the covariates are continuous r.v.'s, as otherwise, the problem of gridding may become even simpler.

To motivate the general q-variate case, we look into the problem of estimating $\Theta(C)$ in (3.7) when C is taken to be a compact set embedded in the essential support of the distribution of the concomitant vector $\underset{\sim}{Z}$. For example, we may take

$$C = \{\underset{\sim}{z} = (z^{(1)},\ldots,z^{(q)})' : a_i \leq z^{(i)} \leq b_i \ , \ i=1,\ldots,q \}, \tag{3.8}$$

where $-\infty < a_i < b_i < +\infty$, for i=1,...,q. Consider then a finite set of m ($= m_1 \ldots m_q$) grid-points $\underset{\sim}{z}(\underset{\sim}{i}) = (z_{i_1}^{(1)},\ldots,z_{i_q}^{(q)})'$, for $i_j = 1,\ldots,m_j$, $j = 1,\ldots,q$, and, we choose these grid-points in such a way that for some given $\varepsilon > 0$,

$$\sup_{\underset{\sim}{z} \in C} \inf_{\underset{\sim}{i}} \rho(\underset{\sim}{z},\underset{\sim}{z}(\underset{\sim}{i})) < \varepsilon . \tag{3.9}$$

In this context, we may take the usual Euclidean norm for the metric $\rho(.)$. Such a design is termed an **ε-affine design** (for a fixed $\varepsilon > 0$). Note that in (3.9) ε is fixed (although it may be chosen small when n is large), so that we have essentially a fixed design. On the other hand, as is usually the case in asymptotic analysis, we may even allow $\varepsilon = \varepsilon_n$ to depend on n in an appropriate manner, so that ε_n is made to converge to 0 as $n \to \infty$. In this case, in order to achieve (3.9) with ε replaced by ε_n, we may need to replace the m_j by a sequence m_{jn} , for j=1,...,q, where the m_{jn} all depend on n and they go to ∞ as $n \to \infty$. Since in this case , $m = m_n$ is a sequence of positive integers, such that $m_n \to \infty$ as $n \to \infty$, we term this a sequential design, although there may not be a major role of any stopping numbers which are usually adopted in the formulation of a sequential design. In the sequential case, in order to validate the statistical analysis with an increasing number (m_n) of grid-points, some additional care has to be made.

In a (fixed) ε-affine design, corresponding to a set of chosen grid-points, the estimates of the $\theta(\underset{\sim}{z}(\underset{\sim}{i}))$ will be shown to be asymptotically independent (for different combinations of the levels $\underset{\sim}{i}$), so that some simplifications can be achieved in the analysis of their asymptotic properties. But, this generally entails some loss of efficiency, mostly due to the nature of the bias terms. The sequential design has a better handle over these bias terms, and although the estimators at the adjacent grid-points may not be asymptotically independent, their inter-dependence structure is of such a simple asymptotic form that a more efficient estimation of $\Theta(C)$ can be contemplated. These constitute the main objective of the current study. We conclude this section with some remarks on the actual construction of such grid-points. In the case of q = 1, the problem is relatively simple: There C is a compact interval [a,b], and we may take $i : a_i = a + 2\varepsilon i$, for i=0,1,...,m, where $m = [(b-a)/2\varepsilon]+ 1$, so that (3.9) holds. The situation with the sequential design is quite analogous, and we need to replace ε by ε_n in the above definition. For $q \geq 2$, the choice of the metric $\rho(.)$ may have a greater impact on this selection of grid points. In practice, choice of rectangular grids is very common. For example, for q= 2, we may divide the compact rectangle C [in (3.8)] into $m_1 m_2$ sub-rectangles whose edges are of length $\sqrt{2}\varepsilon$, so that any point in a sub-rectangle is at most apart from one of the four vertices by ε . In that case, $m_1 = [(b_1-a_1)/\varepsilon\sqrt{2}]+1$, and a similar relation holds for m_2. It is also possible to choose circular grids (or elliptical grids) with appropriate bounds on the diameters so that (3.9) holds.In the choice of C, the scale for the $z^{(1)}$ and $z^{(2)}$ coordinates may not be same and they may also be dependent in a

statistical sense. The rectangular gridding can take care of the inequity of the scales in a natural manner, although it is insensitive to possible dependence of the coordinates. The circular grid is insensitive to both scalar differences and dependence of the coordinates, while the elliptical one can take care of both if the appropriate measures are incorporated in the formulation of the quadratic forms on which the elliptical distances are based. The use of the Mahalanobis distance and the corresponding gridding eliminates this problem, although from robustness considerations, this may not be very desirable. If the concomitant variates are less susceptible to error contaminations and outliers, such robustness considerations are not of utmost importance, and hence, the Mahalanobis metric and gridding may be advocated with greater confidence. The picture is quite similar in the general case of $q \geq 2$. In the concluding section we shall make some further comments on the sequential design case.

4. PRELIMINARY NOTIONS AND REGULARITY ASSUMPTIONS

Let $F(\underset{\sim}{z})$, $\underset{\sim}{z} \in R^q$ be the joint d.f. of the covariate $\underset{\sim}{Z}$. We assume that F admits an absolutely continuous (q-variate) density function $f(\underset{\sim}{z})$, $\underset{\sim}{z} \in R^q$, such that

$$f(\underset{\sim}{z}) > 0 \text{ , for every } \underset{\sim}{z} \in C, \tag{4.1}$$

and the Hessian $(\partial^2/\partial\underset{\sim}{z}\partial\underset{\sim}{z}')f(\underset{\sim}{z})$ exists for all $\underset{\sim}{z} \in C$. Moreover, there exist positive and finite numbers ε , K_0 and α , such that $||\underset{\sim}{z} - \underset{\sim}{z}^*|| < \varepsilon$ implies that

$$||(\partial^2/\partial\underset{\sim}{z}\partial\underset{\sim}{z}')f(\underset{\sim}{z})|_{\underset{\sim}{z}} - (\partial^2/\partial\underset{\sim}{z}\partial\underset{\sim}{z}')f(\underset{\sim}{z})|_{\underset{\sim}{z}^*}|| \leq K_0||\underset{\sim}{z}-\underset{\sim}{z}^*||^{\alpha}, \text{ uniformly in C.} \tag{4.2}$$

For simplicity of presentation, in (4.2), we have used the Euclidean norm , but a very similar formulation holds for a general norm. With respect to the norm (.) in (3.1), we define the D_i as in (3.2), and denote the d.f. and the density function of D by $F_D(d)$ and $f_D(d)$ respectively. Note that the D_i are defined with respect to a pivot $\underset{\sim}{z}_0$ and hence, both $F_D(d)$ and $f_D(d)$ depend on this pivot $\underset{\sim}{z}_0$. We therefore write $f_D(d)$ as $f_D(d;\underset{\sim}{z}_0)$. Note that , by definition,

$$f_D(d;\underset{\sim}{z}_0) = \int_{\{\underset{\sim}{z}:\rho(\underset{\sim}{z},\underset{\sim}{z}_0) = d\}} f(\underset{\sim}{z})d\nu(\underset{\sim}{z};\underset{\sim}{z}_0) \text{ , } \quad d \in R^+ \text{ , } \tag{4.3}$$

where $\nu(\underset{\sim}{z};\underset{\sim}{z}_0)$ stands for the measure on the segment $\{\underset{\sim}{z}: \rho(\underset{\sim}{z},\underset{\sim}{z}_0) = d\}$. For example, if q=1, one has unit mass at the two points z_0 -d and z_0+d, and zero elsewhere. Similarly, for q=2, when $\rho(\underset{\sim}{z},\underset{\sim}{z}_0)$ refers to the Euclidean norm, this measure attaches unit mass to points on the circumference of the circle $||\underset{\sim}{z}-\underset{\sim}{z}_0||^2 = d^2$, and zero elsewhere. For $q \geq 3$, for Euclidean norm, this goes over neatly to that over the sperical surfaces. In any case, we may conclude that both (4.1) and (4.2) hold for $f_D(d;\underset{\sim}{z}_0)$ as well.

As in before, we denote the conditional distribution of Y_i given $\underset{\sim}{Z}_i = \underset{\sim}{z}_0$, by $G(y|\underset{\sim}{z}_0)$ and assume that G(.) has a continuous density $g(y|\underset{\sim}{z}_0)$ for almost all $(y,\underset{\sim}{z}_0)$. Some further differentiability conditions are needed in this context, and these are formulated along the lines of Gangopadhyay and Sen (1993), but extended further to the multivariate case. Note that the conditional d.f. and density of Y_i, given $D_i = d$ (with respect to a pivot $\underset{\sim}{z}_0$), are respectively given by

$$G_D(y \mid \underset{\sim}{z}_0, d) = \{ \int_{\{\underset{\sim}{z}:\rho(\underset{\sim}{z},\underset{\sim}{z}_0)=d\}} G(y|\underset{\sim}{z}) f(\underset{\sim}{z}) d\nu(\underset{\sim}{z};\underset{\sim}{z}_0) \} / f_D(d;\underset{\sim}{z}_0) , \tag{4.4}$$

and

$$g_D(y \mid \underset{\sim}{z}_0, d) = \{ \int_{\{\underset{\sim}{z}:\rho(\underset{\sim}{z},\underset{\sim}{z}_0)=d\}} g(y|\underset{\sim}{z}) f(\underset{\sim}{z}) d\nu(\underset{\sim}{z};\underset{\sim}{z}_0) \} / f_D(d,\underset{\sim}{z}_0) , \tag{4.5}$$

where $\underset{\sim}{z}_0 \in C$ and $y \in R$. We assume that there exists a positive d_0, such that the first and second order partial derivatives of $g_D(y|\underset{\sim}{z}_0,d)$ with respect to d, denoted by $g_D'(.)$ and $g_D''(.)$ respectively, exist for almost all $y,\underset{\sim}{z}_0$ and $d : 0 < d \leq d_0$, and these are dominated by some Lebesgue measurable functions. Moreover,

$$| g_D''(y \mid \underset{\sim}{z}_0, d) - g_D''(y|\underset{\sim}{z}_0, 0) | \leq u_3(y) d^{\alpha} , \tag{4.6}$$

where $u_3(y)$ is Lebesgue measurable, α is a positive number, and (4.6) holds for all $d \leq d_0$. Finally, we assume that

$$I(y \mid \underset{\sim}{z}_0, d) = E[\{(\partial/\partial D) \log g_D(Y \mid \underset{\sim}{z}_0, d) \}^2 \mid D = d] < \infty , \tag{4.7}$$

uniformly in $d : 0 < d \leq d_0$ and $\underset{\sim}{z}_0 \in C$. In passing, we may remark that the Euclidean norm or the one in (3.2) is not scale-equivariant (although the Euclidean one is rotation-invariant). Thus, if the coordinates of $\underset{\sim}{z}$ are on different units of measurements and possibly have different scale parameters, the rationality of a chosen norm has to be made clear; the choice of a norm has a profound effect on (4.7) and it also affects the statistical analysis based on such a choice.

Note that by assumption [in (4.1)], $f(\underset{\sim}{z}) > 0$ for every $\underset{\sim}{z} \in C$. Thus, if we assume that $f(\underset{\sim}{z})$ is (twice) continuously differentiable with respect to the elements of $\underset{\sim}{z}$, then we may write (for $\underset{\sim}{z}$ in a neighborhood of $\underset{\sim}{z}_0$),

$$f(\underset{\sim}{z}) = f(\underset{\sim}{z}_0) + (\underset{\sim}{z} - \underset{\sim}{z}_0)' [(\partial/\partial \underset{\sim}{z}) f(\underset{\sim}{z})]_{\underset{\sim}{z}_0} + \tfrac{1}{2}(\underset{\sim}{z}-\underset{\sim}{z}_0)' \underset{\sim}{A}(\underset{\sim}{z}_0)(\underset{\sim}{z}-\underset{\sim}{z}_0) + o(||\underset{\sim}{z}-\underset{\sim}{z}_0||^2), \tag{4.8}$$

where

$$\underset{\sim}{A}(\underset{\sim}{z}_0) = (\partial^2/\partial \underset{\sim}{z} \partial \underset{\sim}{z}') f(\underset{\sim}{z})|_{\underset{\sim}{z}_0} . \tag{4.9}$$

Thus, whenever $\rho(\underset{\sim}{z},\underset{\sim}{z}_0) = d$ yields a closed contour which is symmetric around $\underset{\sim}{z}_0$ (in a well defined manner), we have from (4.3) and (4.8) that for $d \to 0$,

$$f_D^0(d) = f(\underset{\sim}{z}_0) a(d) + (1/2) Q_1(d) + o(d^2 a(d)) \text{ , as } d \to 0, \tag{4.10}$$

where a(d) is the surface area of the countour $\{\underset{\sim}{z}: \rho(\underset{\sim}{z},\underset{\sim}{z}_0) = d\}$ and

$$Q_1(d) = \int_{\{\underset{\sim}{z}:\rho(\underset{\sim}{z},\underset{\sim}{z}_0)=d\}} (\underset{\sim}{z}-\underset{\sim}{z}_0)' \underset{\sim}{A}(\underset{\sim}{z}_0)(\underset{\sim}{z}-\underset{\sim}{z}_0) \, d\nu(\underset{\sim}{z};\underset{\sim}{z}_0) . \tag{4.11}$$

Note that by the Courant theorem on quadratic forms, $(\underset{\sim}{z}-\underset{\sim}{z}_0)' \underset{\sim}{A}(\underset{\sim}{z}_0)(\underset{\sim}{z}-\underset{\sim}{z}_0)/||\underset{\sim}{z}-\underset{\sim}{z}_0||^2 \leq a^*(\underset{\sim}{z}_0)$= largest characteristic root of $\underset{\sim}{A}(\underset{\sim}{z}_0)$, so that by (4.11), we obtain that $Q_1(d)/a(d) = O(d^2)$, as $d \to 0$. At this stage, it may be more pertinent to make some comments on a(d) and $Q_1(d)$. First, in the case of q=1, a(d) = 2 and $Q_1(d) = 2cd^2$ where $c = f''(z_0)$. For q =2, $a(d) = 2\pi d$ when we use the Euclidean norm for $\rho(.)$, and $Q_1(d)$ depends on the Hessian $\underset{\sim}{A}(\underset{\sim}{z}_0)$; also, $Q_1(d) = O(d^3)$. In the higher dimensional case, with the Euclidean norm, we may set a(d) as the surface area of a sphere of radius d , so that it is $O(d^{q-1})$, and verify that $Q_1(d) = O(d^{q+1})$. Secondly, both a(d) and $Q_1(d)$ may depend on the particular norm $\rho(.)$ used for the definition of the D_i. It may be quite appropriate to set a condition that as $d \to 0$, $\rho(\underset{\sim}{z},\underset{\sim}{z}_0)/||\underset{\sim}{z}-\underset{\sim}{z}_0||$ for

$||\underset{\sim}{z}-\underset{\sim}{z}_0|| = d$, converges to a positive limit. In that case, a locally quadratic approximation for $\rho(.)$ works out well, and this simplifies the asymptotics considerably. Finally, since the Hessian $\underset{\sim}{A}(\underset{\sim}{z}_0)$ is unknown, the explicit form for $Q_1(d)$, even for $d \downarrow 0$, is not known, and hence, as in the usual nearest neighborhood methodology, there may be a bias term for the estimators of the $\theta(\underset{\sim}{z}_0)$ which is generally of the same order of magnitude as their standard error. As such, we need to examine this bias and we shall discuss more about this later on.

For the conditional d.f. $G(y|\underset{\sim}{z})$, we make a similar expansion around $\underset{\sim}{z}_0$, and obtain that for small $||\underset{\sim}{z} - \underset{\sim}{z}_0||$,

$$G(y|\underset{\sim}{z}) = G(y|\underset{\sim}{z}_0) + (\underset{\sim}{z}-\underset{\sim}{z}_0)'[(\partial/\partial\underset{\sim}{z})G(y|\underset{\sim}{z})]_{\underset{\sim}{z}_0} + (1/2)(\underset{\sim}{z}-\underset{\sim}{z}_0)'\underset{\sim}{B}(\underset{\sim}{z}_0)(\underset{\sim}{z}-\underset{\sim}{z}_0) + o(||\underset{\sim}{z}-\underset{\sim}{z}_0||^2), \tag{4.12}$$

where $\underset{\sim}{B}(\underset{\sim}{z}_0)$ is the Hessian of $G(y|\underset{\sim}{z})$ with respect to $\underset{\sim}{z}$ at $\underset{\sim}{z}_0$. Let us also denote by

$$\underset{\sim}{C}(\underset{\sim}{z}_0) = [(\partial/\partial\underset{\sim}{z})G(y|\underset{\sim}{z})]_{\underset{\sim}{z}_0}[(\partial/\partial\underset{\sim}{z}')f(\underset{\sim}{z})]_{\underset{\sim}{z}_0}, \tag{4.13}$$

and note that both $\underset{\sim}{B}(\underset{\sim}{z}_0)$ and $\underset{\sim}{C}(\underset{\sim}{z}_0)$ are possibly dependent on y as well. Then, from (4.4), (4.9), (4.12) and (4.13), we have for small d,

$$f_D(d;\underset{\sim}{z}_0)G_D(y|\underset{\sim}{z}_0,d) = f(\underset{\sim}{z}_0)G(y|\underset{\sim}{z}_0)\,a(d) + (1/2)G(y|\underset{\sim}{z}_0)Q_1(d) + (1/2)f(\underset{\sim}{z}_0)Q_2(d) + Q_3(d) + o(Q_1(d)+Q_2(d)+Q_3(d)), \tag{4.14}$$

where $Q_2(d)$ and $Q_3(d)$ are defined as in (4.11) with $\underset{\sim}{A}(\underset{\sim}{z}_0)$ being replaced by $\underset{\sim}{B}(\underset{\sim}{z}_0)$ and $\underset{\sim}{C}(\underset{\sim}{z}_0)$ respectively. Therefore, using (4.11) and (4.14), and integrating over $d: 0 \le d \le r$, for small r, we have for the average conditional d.f.,

$$\begin{aligned} G_r(y|\underset{\sim}{z}_0) &= \{\int_0^r f_D(d;\underset{\sim}{z}_0)G_D(y|\underset{\sim}{z}_0,d)dd\}/\{\int_0^r f_D(d;\underset{\sim}{z}_0)dd\} \\ &= \frac{f(\underset{\sim}{z}_0)G(y|\underset{\sim}{z}_0)v(r) + \frac{1}{2}G(y|\underset{\sim}{z}_0)Q_1^*(r) + \frac{1}{2}f(\underset{\sim}{z}_0)Q_2^*(r) + Q_3^*(r) + o(Q_1^*(r))}{f(\underset{\sim}{z}_0)v(r) + (1/2)Q_1^*(r) + o(Q_1^*(r))} \\ &= G(y|\underset{\sim}{z}_0) + \{f(\underset{\sim}{z}_0)Q_2^*(r) + 2Q_3^*(r)\}/\{2f(\underset{\sim}{z}_0)v(r)+Q_1^*(r)\} + o((Q_1^*(r))/v(r)), \end{aligned} \tag{4.15}$$

where

$$v(r) = \int_0^r a(d)dd \quad \text{and} \quad Q_j^*(r) = \int_0^r Q_j(d)dd,\ j=1,2,3, \text{ for } r > 0. \tag{4.16}$$

In a similar manner, in (4.12), we replace $G(y|\underset{\sim}{z})$ by the density $g(y|\underset{\sim}{z})$, and carry out the analyis as in (4.13) through (4.16). This will lead us to a parallel expression for $g_r(y|\underset{\sim}{z}_0)$ in terms of $g(y|\underset{\sim}{z}_0)$, $f(\underset{\sim}{z}_0)$ and $v(r)$ as well as analogues of the $Q_j^*(r)$, $j=1,2,3$. When z is real valued, $v(r) = 2r$ while the $Q_j^*(r)$ are all $O(r^3)$, so that in (4.15), the remainder term is $o(r^2)$. However, when $\underset{\sim}{z}$ is vector valued, although this remainder term remains as $o(r^2)$, the second term on the right hand side of (4.15) is generally much more complex. Nevertheless, noting that the $Q_j^*(r)$ and $v(r)$ all converge to 0 as $r \to 0$, we may apply the L'Hopital's rule and obtain that as $r \to 0$,

$$\{f(\underset{\sim}{z}_0)Q_2^*(r) + 2Q_3^*(r)\}/\{2f(\underset{\sim}{z}_0)v(r) + Q_1^*(r)\} = r^2h(y,\underset{\sim}{z}_0) + o(r^2), \tag{4.17}$$

where $h(y,\underset{\sim}{z}_0)$ depends on the gradients and the Hessians of $f(\underset{\sim}{z})$ and $G(y|\underset{\sim}{z})$ at $\underset{\sim}{z}_0$, and it may be of quite involved form. Thus, for $r \to 0$, we may write the right hand side of (4.15) as

$$G_r(y|\underset{\sim}{z}_0) = G(y|\underset{\sim}{z}_0) + r^2 h(y,\underset{\sim}{z}_0) + o(r^2). \tag{4.18}$$

In a similar manner, we have

$$g_r(y|\underset{\sim}{z}_0) = g(y|\underset{\sim}{z}_0) + r^2 h^*(y,\underset{\sim}{z}_0) + o(r^2) \text{ , as } r \to 0, \tag{4.19}$$

where $h^*(y,\underset{\sim}{z}_0)$ depends on the gradients and Hessians of $f(\underset{\sim}{z})$ and $g(y|\underset{\sim}{z})$ at $\underset{\sim}{z}_0$. For the case of real valued z, Gangopadhyay and Sen (1992,1993) have obtained similar expressions for the average conditional d.f. with a more specific formulation of r (viz., $r = r_n = O(n^{-1/5})$. However, in the case of vector-valued concomitant variates, the rate of r_n is dependent on q, the number of components, and hence, the current formulation makes it much more flexible to adopt such a representation for uselful statistical analysis. In this context, we may note that for the Euclidean norm [refer to (3.1)] , corresponding to a value of d (> 0), the volume of the sphere $\{\underset{\sim}{z}: \rho(\underset{\sim}{z},\underset{\sim}{z}_0) \leq d\}$ is $O(d^q)$, for $q \geq 1$, so that we have $v(r) = O(r^q)$ as $r \to 0$. A similar order holds for general norms. On the other hand, if we choose a nearest neighborhood of k_n members close to the point $\underset{\sim}{z}_0$, then under (4.1), we have

$$n^{-1}k_n = O(r_n^q) \text{ when } k_n/n \to 0 \text{ as } n \to \infty. \tag{4.20}$$

As such, if we let $r_n \sim n^{-\lambda}$, for some $\lambda > 0$, we may set

$$k_n \sim n^{1-q\lambda} \text{ , as } n \to \infty, \tag{4.21}$$

and this naturally put an upper bound for λ , namely, $1/q$. Moreover, as we shall see later on that the bias term arising in this conditional quantile estimation, as can be dictated by (4.18), is $O(r_n^2)$, so that the contribution of the bias term in the asymptotic mean square error of the estimator is $O(r_n^4) = O(n^{-4\lambda})$, while the variance term is $O(k_n^{-1}) = O(n^{q\lambda-1})$. Hence, we should select λ in such a way that

$$1 - q\lambda \leq 4\lambda \text{ i.e., } \lambda \leq 1/(q+4) , \tag{4.22}$$

so that

$$k_n = O(n^{4/(q+4)}) \text{ , as } n \to \infty . \tag{4.23}$$

If we choose $k_n = o(n^{4/(q+4)})$, the contribution of the bias term in the asymptotic mean square error will be negligible, but the order of the mean square error will be $o(n^{-4/(q+4)})$, so that we have a somewhat slower rate of convergence. We will discuss more about this problem in the next section. We conclude this section with a remark that since C is taken as a compact set and $\underset{\sim}{z}_0 \in C$, the discussion made above pertains to the whole domain C and the technicalities will be made clear in the next section. Further, the current findings also pertain to the kernel method of smoothing for conditional quantiles in the case of vector valued concomitant variates. For the latter approach, one may then need to choose an appropriate multivariate kernel density where scale-equivariance and inter-dependence may need to be taken into account in a proper way. While this is no problem, it generally induces lack of robustness, and moreover, for $q \geq 2$, the asymptotic results provide reasonable approximations only for very large sample sizes. In view of the essential similarity in the two

approaches (and parallel results with respect to their asymptotics), we drop the details to avoid essential repetitions of similar analysis. We may refer to Bhattacharya and Gangopadhyay (1990) for some interesting results on such asymptotic equivalence in the case of $q = 1$, and these remain in tact for higher dimensional concomitant variates too.

5. STATISTICAL ANALYSIS FOR ε-AFFINE DESIGNS

We define an ε-affine design as in (3.8)-(3.9). In such a case, we have a fixed number m ($=m_\varepsilon$) of gridpoints which we denote by

$$\underset{\sim}{z}(\underset{\sim}{i}) = (z_{i_1}^{(1)},\ldots,z_{i_q}^{(q)})' \ , \ \underset{\sim}{i} = (i_1,\ldots,i_q)' \in I \ , \tag{5.1}$$

where

$$I = \{\underset{\sim}{i} : i_j=1,\ldots,m_j \ , \ j=1,\ldots,q \} \ ; \ m = m_1\ldots m_q. \tag{5.2}$$

Corresponding to each $\underset{\sim}{i} \in I$, we consider the covariate point $\underset{\sim}{z}(\underset{\sim}{i})$ as defined above, and choosing $\underset{\sim}{z}(\underset{\sim}{i})$ as the pivot, we proceed to estimate the conditional quantile $\theta(\underset{\sim}{z}(\underset{\sim}{i})) = \theta(\underset{\sim}{i})$, say. We denote this set of estimators by

$$\hat{\Theta}_n(C, \varepsilon) = \{ \ \hat{\theta}_n(\underset{\sim}{i}) : \underset{\sim}{i} \in I \}, \tag{5.3}$$

where , for each $\underset{\sim}{i} \in I$, the estimator $\hat{\theta}_n(\underset{\sim}{i})$ is defined as in (3.6) with the pivot $\underset{\sim}{z}_o$ being replaced by $\underset{\sim}{z}(\underset{\sim}{i})$. These fixed-point estimators are then to be incorporated in providing a smooth estimator of $\underset{\sim}{\theta}(\underset{\sim}{z})$, $\underset{\sim}{z} \in C$.

For the case of $q = 1$, i.e., real valued z, and a single point z_o, asymptotic properties of this conditional quantile estimator have been studied in detail by Bhattacharya and Gangopadhyay (1990) and Gangopadhyay and Sen (1992,1993), among others. For $q = 1$, we have $k_n = O(n^{4/5})$, and the following representation for the estimator:

For k_n lying in (k_{no}, k_{n1}) where $k_{no} = [an^{4/5}]$, $k_{n1} = [bn^{4/5}]$ and $0 < a < b < \infty$, as n increases,

$$\hat{\theta}_{n,k_n}(z_o) - \theta(z_o) = \beta(z_o)(n^{-1}k_n)^2 + (k_n g(\theta(z_o)|z_o))^{-1}\Sigma_{i=1}^{k_n} [I(Y_{ni}^* > \theta(z_o)) - (1-p)] + R_{n,k_n} \ , \tag{5.4}$$

where

$$\max\{ |R_{n,k_n}| : k_{no} \le k_n \le k_{n1} \} = O(n^{-3/5}\log n) \text{ a.s.}, \tag{5.5}$$

and letting $G(.)$ for $G(y|z_o)$ and $G_D(.)$ for $G_D(y|z_o,d)$, the Y_{ni}^* are conditionally independent r.v.'s defined by

$$Y_{ni}^* = G^{-1} o\, G_D(Y_{S_i}|Z_{S_i}) \ , \ i=1,\ldots,n; \tag{5.6}$$

finally, the bias term $\beta(z_o)$ depends on $\theta(z_o)$, z_o and on the associated densities.

In this respect, the contiguity-based proof of (5.4) provided by Gangopadhyay and Sen (1992) is a further simplification of the original proof in Bhattacharya and Gangopadhyay (1990). Another important observation due to Bhattacharya and Gangopadhyay (1990) is the incorporation of all values of $k_n \in (k_{no},k_{n1})$ to eliminate the bias term in (5.4) with a small increase in the asymptotic variance of the estimator. This feature is of considerable importance for the subsequent analysis in the current study.

For a general q (≥ 1), we have noticed in (4.23) that an optimal k_n (for a fixed point $\underset{\sim}{z}_o$), achieving a balance between the bias and the sampling error of the estimator, is given by $k_n = O(n^{4/(q+4)})$. With such a choice of k_n, one can readily extend the contiguity results in Gangopadhyay and Sen (1992, 1993) to the q-variate case, and, as such, using (4.18)-(4.19) along with some analysis very similar to that in the univariate case, we arrive at the following :

For k_n belonging to $\{k_{no} = [an^{4/(q+4)}],\ k_{n1} = [bn^{4/(q+4)}]\}$, for $0 < a < b < \infty$, the asymptotic representation in (5.4) holds for the general case too, where in (5.5), the right hand side is $O(n^{-3/(q+4)} \log n)$ a.s., as $n \to \infty$, and $\beta(\underset{\sim}{z}_o)$ is a more involved function of the joint density $f(\underset{\sim}{z}_o)$ and the conditional d.f. $G(.|\ \underset{\sim}{z}_o)$.

As in Bhattacharya and Gangopadhyay (1990), we may then incorporate the set of estimators $\{\ \hat{\theta}_{n,k_n}(\underset{\sim}{z}_o);\ k_{no} \leq k_n \leq k_{n1}\ \}$ in eliminating the leading bias term in (5.4) at the cost of a small increase in the asymptotic variance function. Towards this objective, we consider a linear combination :

$$\Sigma_{k=k_{no}}^{k_{n1}} a_{nk}\, \hat{\theta}_{n,k}(\underset{\sim}{z}_o) = \hat{\theta}^*_{n,k_n}(\underset{\sim}{z}_o) \ , \text{say.} \tag{5.7}$$

Note that the a_{nk} need not all be nonnegative, but they must add upto 1, and moreover, by using (5.4), we have the restraint:

$$\Sigma_{k=k_{no}}^{k_{n1}} (n^{-1}k)^2\, a_{n,k} = 0 \ . \tag{5.8}$$

Finally, for the linear combination in (5.7), the variance function cropping up from the principal term on the right hand side of (5.4) is equal to

$$[p(1-p)][g(\ \theta(\underset{\sim}{z}_o)\ |\underset{\sim}{z}_o)]^{-2} \Sigma_{k=k_{no}}^{k_{n1}} \Sigma_{k'=k_{no}}^{k_{n1}} \{(k \wedge k')/kk'\}\, a_{n,k} a_{n,k'} \ . \tag{5.9}$$

Therefore, the basic task is to minimize (5.9) with respect to the $a_{n,k}$ subject to the restrain in (5.8) as well as the additivity (to unity) restraint. In this sense, the problem is isomorphic to the univariate case [treated in Bhattacharya and Gangopadhyay (1990)] and hence, the same solution applies here. In this setup, with respect to the choice of k_{no} and k_{n1}, the constants a and b are to be so chosen that a/b is small, and b is large. Of course, this has to be made keeping in mind that k_{n1} is to be $o(n)$ (and $O(n^{4/(q+4)})$).

Let us now go back to the grid-points $\underset{\sim}{z}(\underset{\sim}{i})$, defined by (5.1) and (5.2). For each $\underset{\sim}{z}(\underset{\sim}{i})$, $\underset{\sim}{i} \in I$, we proceed as in above, and obtain the bias-reduced asymptotically optimal estimator $\hat{\theta}^*_{n,k_n}(\underset{\sim}{z}(\underset{\sim}{i}))$. Thus, corresponding to the set of m grid-points, $\underset{\sim}{z}(\underset{\sim}{i})$, $\underset{\sim}{i} \in I$, we obtain the set of estimators:

$$\{\ \hat{\theta}^*_{n,k_n}(\underset{\sim}{z}(\underset{\sim}{i}))\ ;\ \underset{\sim}{i} \in I\ \} \ . \tag{5.10}$$

For the case of real valued z, Gangopadhyay and Sen (1993) have studied the asymptotic behavior of the conditional quantile process by allowing z to vary over a (compact) interval . It was shown there that for two distinct points, say, z_0 and z^0 , as n increases, the intersection of the induced order statistics sets for the two neighborhoods becomes asymptotically a null set, and hence, the two estimators are asymptotically stochastically

independent. They used the contiguity approach to explore this point-wise independence behavior, and that approach extends readily for the general case of q-variate $\underset{\sim}{z}$'s, for any (fixed) $q \geq 1$, whenever, k_n is chosen accordingly [viz., (4.23)]. If we define k_{no} and k_{n1} as in before, we may note that this phenomenon of asymptotic null intersections of the two sets of induced order statistics holds, and hence, the set of adjusted estimators in (5.10) have asymptotically independent elements, so that their asymptotic marginal d.f,'s dictate the m-variate joint d.f. As such, we may conclude that as n increases,

$$n^{2/(q+4)}\{ \hat{\theta}^{*}_{n,k_n}(\underset{\sim}{z}(i)) - \theta(\underset{\sim}{z}(i)) , \underset{\sim}{i} \in I \} \xrightarrow{\mathcal{D}} N_m(\underset{\sim}{0}, \underset{\sim}{D}_m), \tag{5.11}$$

where $\underset{\sim}{D}_m$ is a diagonal matrix with the elements given by the asymptotic variances of the marginal estimators. The matrix $\underset{\sim}{D}_m$ depends not only on the conditional densities $g(.|\underset{\sim}{z})$ at these values of $\underset{\sim}{z} = \underset{\sim}{z}(i)$, $\underset{\sim}{i} \in I$, but also on other functionals depending on the joint density $f(\underset{\sim}{z}(i))$ and other derivatives of f(.) and g(.) in a rather involved manner. In order to provide a (simultaneous) confidence interval for the discrete set $\theta(\underset{\sim}{z}(i))$; $\underset{\sim}{i} \in I$, one needs to have the knowledge of the elements of $\underset{\sim}{D}_m$. This brings us to the problem of estimating $\underset{\sim}{D}_m$ in a consistent manner. Since, we are basically dealing with quantiles (and moreover in a conditional setup), the classical jackknife method of variance estimation may not work out that well in the current setup. Although it may be quite possible to use delete-k jackknifing for (moderately) large values of d (small compared to k_n), we shall find it convenient to use the bootstrap method. For the real valued covariate case, at a single point z_o , Gangopadhyay and Sen (1990) studied bootstrap confidence intervals for $\theta(z_o)$ wherein bootsrapping was incorporated through the raw estimators $\hat{\theta}_{n,k_n}(z_o)$.But their Theorems 2.5, 2.6, 2.7 and 2.8 contain all the ingredients for the successful adoptation of bootstrapping for the adjusted estimator $\hat{\theta}^{*}_{n,k_n}(z_o)$ as well. Omitting the details of some manipulations, we may say at this stage that this bootstrap methodology works out as well for the q-variate $\underset{\sim}{z}_o$ when k_n is adjusted accordingly [see (4.23)]. Thus, for each $\underset{\sim}{z}(i)$, $\underset{\sim}{i} \in I$, we may use this bootstrap methodology to provide a consistent estimator of the corresponding diagonal element of $\underset{\sim}{D}_m$. Since m is a fixed positive integer, the consistency of the estimator of $\underset{\sim}{D}_m$ follows from that of the individual diagonal elements. We intend to present these technicalities in a somewhat more general setup in a future communication.

We may remark in passing that if $\{\hat{d}(\underset{\sim}{i})\}^2$ is the estimator of the i-th diagonal element of $\underset{\sim}{D}_m$ (as obtained by the bootstrap method), then, from (5.11), we have

$$\{ n^{2/(q+4)}\{ \hat{\theta}_{n,k_n}(\underset{\sim}{z}(i)) - \theta(\underset{\sim}{z}(i))\}/\hat{d}(\underset{\sim}{i}) , \underset{\sim}{i} \in I \} \xrightarrow{\mathcal{D}} N_m(\underset{\sim}{0}, \underset{\sim}{I}_m), \tag{5.12}$$

which may be used to draw a simultaneous confidence interval for the $\theta(\underset{\sim}{z}(i))$. It may be naturally tempting to extend this result to cover the entire $\theta(\underset{\sim}{z})$, $\underset{\sim}{z} \in C$. However, there are some hidden points which merit particular attention. First, the $\underset{\sim}{z}(i)$, $\underset{\sim}{i} \in I$ and m are fixed for a fixed design. Thus, if we want to interpolate $\theta(\underset{\sim}{z})$ from its neighboring $\theta(\underset{\sim}{z}(i))$

there may be a bias unless the $\theta(\underset{\sim}{z})$ is itself a polynomial of a finite degree (in the sense that its partial derivatives (with respect to $\underset{\sim}{z}$) of order r+1 or more are all equal to 0, for some positive integer r). Although under (3.9), such a bias term will be very small (i.e., o(1)), but the presence of the scaling factor $n^{2/(q+4)}$ in (5.11) or (5.12) will make the contribution unbounded for large n. Secondly, if we only assume some locally smoothness conditions on $\theta(\underset{\sim}{z})$, then, we need to allow ε to depend on n as well, and this will lead to $m = m_n \to \infty$ as $n \to \infty$. We shall treat this in the next section. Thirdly, in weak convergence results, convergence of finite-dimensional distributions may not yield the same for the entire stochastic process, and one needs to verify the 'tightness' or compactness condition to bridge in this gap. Since we are faced with a multi-parameter stochastic process here (viz., $n^{2/(q+4)} \{\hat{\theta}^*_{n,k_n}(\underset{\sim}{z}) - \theta(\underset{\sim}{z})\}$, $\underset{\sim}{z} \in C$), verification of such a tightness condition may require more delicate analysis than given in Billingsley (1968), and we relegate this to a future study. We conclude this section with a positive note on the utility of this ε-affine design in practice. In many situations, although the covariates are continuous r.v.'s, for all practical purposes, they are generally recorded on interval-scales (although the width of the intervals may be small depending on the precision of the recording devise). In such a case, corresponding to the mid-points of these intervals, a finite system of grid-points may be constructed, and for the given set, (5.11) or (5.12) may be incorporated to draw appropriate statistical conclusions. As we need to stick to the grid-points mostly for predictive purposes , there is no need to go over to the continuous case. However, the choice of ε has to be made accordingly.

6. SOME REMARKS ON SEQUENTIAL DESIGNS

If the $\theta(\underset{\sim}{z})$, $\underset{\sim}{z} \in C$ were of parametric form, i.e., $\theta(\underset{\sim}{z})$ can be described in terms of a (possibly nonlinear) function of $\underset{\sim}{z}$ involving only a finite number of algebraic constants (parameters), then one may use a fixed design for drawing conclusions from the given data set. This pleasant feature vanishes when $\theta(\underset{\sim}{z})$ is treated to be fully of a nonparametric form. In such a case, a mesh for $\underset{\sim}{z}$, no matter how dense it is, may not suffice, and one way of eliminating this problem is to take recourse to a sequential designs where the diameter of the mesh is made to depend on the sample size n in a way that the contribution of the bias term in the asymptotic mean squared error can be controlled if not eliminated to a greater extent. Since not much work has been done in this direction even for the case of q = 1 (i.e., real valued z), we present here an overview of the univariate case. The general case of q-variate $\underset{\sim}{Z}$'s will be treated in a more elaborate manner in a subsequent communication.

When specialized to the case of q = 1, we have $k_n = O(n^{4/5})$. In this case, for a fixed z_0, we may proceed as in (5.7) through (5.10), and consider the adjusted (Bhattacharya-Gangopadhyay) estimator $\hat{\theta}^*_{n,k_n}(z_0)$, as defined by (5.10). The main advantage of using this estimator, instead of a single $\hat{\theta}_{n,k_n}$ with an optimal choice of k_n , is that the latter may depend very much on the pivot z_0, and hence, when z_0 is allowed to vary over C, it may be

highly variable. Besides, such optimal k_n are to be estimated from the sample data, and this may call for even larger sample sizes for validity of the proposed procedure.Further, the asymptotic elimination of the bias term for the estimator in (5.10) constitutes another attractive feature of this estimator.

In this case of q=1, C is a closed interval, say, $[\alpha, \beta]$ where $-\infty < \alpha < \beta < \infty$. We may consider a partition of C into m_n sub-intervals $C_{nj} = [\alpha+j.h_n, \alpha+(j+1)h_n]$, where $h_n = (\beta-\alpha)/m_n$, for $j = 0,\ldots, m_n-1$. Note that the length of C_{nj} is equal to $h_n = O(m_n^{-1})$. Next, we choose m_n in such a way that $m_n k_n/n \to 0$ but $m_n^2 k_n/n \to \infty$ as $n \to \infty$. Typically, we may let $m_n \sim n^w$ for some w: $1/10 < w < 1/5$. Consider then the set of points $z(i) = \alpha + i m_n^{-1}(\beta-\alpha)$, for $i=0,1,\ldots,m_n$. It is clear from the representation in (5.4) and the reconstruction in (5.10) that the set of adjusted estimators

$$\hat{\theta}^*_{n,k_n}(z(i)) \ , \ i= 0,1,\ldots,m_n \tag{6.1}$$

consists of asymptotically (stochastically) independent coordinates, and in view of (5.5), it can be shown that the remainder terms for the coordinatewise representations for the elements in (6.1) are uniformly $o(n^{-2/5})$ a.s., as $n \to \infty$. As such, (5.11) extends directly to this increasing dimensional context, and this in turn provides us with the necessary tools for carrying out more refined asymptotic analysis. [For notational simplicity, we have suppressed the subscript n in z(i), but it will be understood.]

We adopt the local smoothing scheme which is classically related to the moving average scheme in time series analysis. For a given $t \in C = [\alpha,\beta]$, we identify the subinterval C_{nj} where t belongs to. As such, we identify the nearest r grid-points z(i), such that $|z(i) - t|$ values are the smallest r order statistics among the m_n values of $|z(i)-t|$, $i=1,\ldots,m_n$. In this context, we may choose r to be a fixed positive integer (so as to accommodate a polynomial smoother of degree r-1), or even, we may allow r to be dependent on n in a way that $r_n/m_n \to 0$ as $n \to \infty$. For simplicity of presentation, we consider only the case of a fixed r (≥ 1). Denote by $S_{nr}(t)$ the set of r of the z(i)'s belonging to this nearest (r) neighborhood of t, and note that the subset of the elements in (6.1) belonging to this set consiststs of asymptotically independent elements, and moreover the corresponding $\theta(z(i))$, $i \in S_{nr}(t)$, are locally packed, so that a smooting plan may as well be adopted for them. Basically, if $z(i_o)$ is the centroid of $S_{nr}(t)$, we set

$$\theta(t) = \theta(z(i_o)) + (t- z(i_o))\ \gamma_1 + \ldots + (t- z(i_o))^s\ \gamma_s \ , \tag{6.2}$$

where s is a positive integer , not larger than r-1, and the γ_j are the coefficients of the various power terms in this polynomial expansion. It is also possible to replace the series in (6.2) in terms of orthogonal polynomials, but the modifications are trivial, and hence, we shall not enter into the detailed deliberations. For each of the estimates $\hat{\theta}^*_{n,k_n}(z(i))$, $i \in S_{nr}(t)$, for their population counterparts, we use (6.2) with t = z(i), and as these estimates are asymptotically independent, and as their local packedness ensure the closeness of their asymptotic variances, we are in a position to use the classical least squares method to estimate $\gamma_1,\ldots, \gamma_s$ from this set exclusively. We may remark

in passing that in (6.2), we allow the possibility of variation of the γ_j from one bucket $S_{nr}(t)$ to another $S_{nr}(t')$, so that these least squares estimators are to be computed for each possible bucket separately; there are in all m_n such buckets (as for each C_{nj} , there will be two buckets (with only one element not in common) depending on whether t is closer to its lower or upper limit, for $j=0,\ldots,m_n-1$). Also, along with the estimates of the γ_j we get the least squares estimator of $\theta(z(i_o))$ too, and when all these estimators are plugged in (6.2), we obtain a smooth estimator of $\theta(t)$, for $t \in C_{nj}$. In this context, we may note that in order that the centroid $z(i_o)$ is itself a grid-point , we take r to be an odd positive integer. This will entail that $\Sigma_{i\in S_{nr}(t)} z(i)/r = z(i_o)$. Secondly, we note that $|t - z(i_o)| = O(m_n^{-1})$, so that the contributions of the second and higher order terms in (6.2) are $O(m_n^{-2})$, which by the very choice of m_n is $o(n^{-1/5})$. Hence, their contributions are asymptotically negligible even when we normalize the estimator by the scale factor $n^{2/5}$. This shows that we could have taken a locally linear function in (6.2) without affecting the asymptotic results, but the inclusion of these higher oreder terms would be more appropriate for moderate sample asymptotics . Let us denote the local least squares estimators of $\theta(z(i_o))$ and $\gamma_1,\ldots,\gamma_s$ in (6.2) by $\ddot{\theta}_n(z(i_o))$ and $\ddot{\gamma}_{n1},\ldots,\ddot{\gamma}_{ns}$, respectively. As such, we obtain that

$$n^{2/5}(\ddot{\theta}_n(t) - \theta(t)) = n^{2/5}(\ddot{\theta}_n(z(i_o)) - \theta(z(i_o))) + (t - z(i_o))n^{2/5}(\ddot{\gamma}_{n1} - \gamma_1) + o_p(n^{-1/5}), \tag{6.3}$$

where by virtue of the adopted design, it can be shown that $n^{2/5}(\ddot{\theta}_n(z(i_o)) - \theta(z(i_o)))$ and $(r/m_n)n^{2/5}(\ddot{\gamma}_{1n} - \gamma_1)$ are asymptotically independent and normally distributed with means 0 and finite variances, while $t - z(i_o) = O(m_n^{-1})$. Therefore, from (6.3), we conclude that as n increases,

$$n^{2/5}(\ddot{\theta}_n(t) - \theta(t)) \xrightarrow{\mathcal{D}} N(0 , \sigma^2(z(i_o))\cdot r^{-1}[1 + m_n^2(t - z(i_o))^2 12/(r^2-1)]) \tag{6.4}$$

where

$$\sigma^2(z(i_o)) = \text{Asymp. Var. of } n^{2/5}(\hat{\theta}^*_{n,k_n}(z(i_o)) - \theta(z(i_o))), \tag{6.5}$$

and $m_n^2(t - z(i_o))^2 \le (\beta - \alpha)^2/4$, for every $t \in C$. This point-wise asymptitic normality results extends to the entire interval $C = [\alpha,\beta]$, and therefore the weak convergence of $W_n = \{ n^{2/5}(\ddot{\theta}_n(t) - \theta(t)), t \in C \}$ to a Gaussian process follows readily.In this respect, (6.1) and the representation in (6.3) provide the necessary tools for the verification of the compactness condition. However, we may note that the $\ddot{\theta}_n(z(i_o))$ are all not asymptotically independent , but whenever $|z(i_o) - z(i_o')| > [2r+1)](\beta - \alpha)/m_n$, they are so. This induces a local dependence in W_n when $|t - t'| = O(r/m_n)$, and also suggests that r should not be unnecessarily chosen large. On the other hand, point-wise asymptotic independence of $W_n(t)$ holds, although the variance function $\sigma^2(t)$, $t \in C$ is generally not stationary. Pointwise bootstrapping for $\sigma^2(t)$ can be made along the same line as considered before, and hence, can be used to make suitable confidence statement about $\theta(t)$. For this purpose too, we may confine ourselves to the basic elements in (6.1) and obtain the bootstrap

estimators of the corresponding $\sigma^2(z(i))$; we denote these by $\hat{\sigma}^{*2}(z(i))$, $i=0,\ldots,m_n$. Then by (6.4) and the stochastic convergence of the $\hat{\sigma}^{*2}(z(i))$, we claim that as n increases,

$$\{ n^{2/5}[\ \hat{\theta}_n(z(i)) - \theta(z(i))\ ]/\hat{\sigma}^*(z(i)) ,\ i = 0,1,\ldots,m_n\} \xrightarrow{\mathcal{D}} N(\underset{\sim}{0},\ \underset{\sim}{I}_{m_n}). \quad (6.6)$$

Since m_n is increasing in n (but slowely enough), we may take advantage of the convergence properties of normal extreme values, and conclude that as n increases ,

$$P\{ \max_{0\le i\le m_n} n^{2/5} \mid \hat{\theta}_n(z(i)) - \theta(z(i))\mid /\hat{\sigma}^*(z(i)) \le (2\log m_n)^{1/2} \} \to 1, \quad (6.7)$$

so that a simultaneous coverage predictive region for $\theta(t)$ can be obtained from (6.7).

ACKNOWLEDGEMENTS.

The author is grateful to the referee for his critical reading of the manuscript and valuable comments on it.

References.

1. Bhattacharya, P.K. (1974). Convergence of sample paths of normalized sums of induced order statistics. Ann. Statist. 2, 1034-1039.

2. Bhattacharya, P.K. and Gangopadhyay, A.K. (1990). Kernel and nearest neighbor estimation of a conditional quantile. Ann. Statist. 17, 1400-1415.

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

4. Gangopadhyay, A.K. and Sen, P.K. (1990). Bootstrap confidence intervals for conditional quantile functions. Sankhya, Ser.A 52, 346-363.

5. Gangopadhyay, A.K. and Sen, P.K. (1992). Contiguity in nonparametric estimation of a conditional functional. in Nonparametric Statistics and Related Topics (ed.A. Saleh), North Holland, Amsterdam, pp.141-162.

6. Gangopadhyay, A.K. and Sen, P.K. (1993). Contiguity in Bahadur-type representations of a conditional quantile and application in conditional quantile process. In Statistics and Probability, a Raghu Raj Bahadur Festschrift (eds:J.K.Ghosh et al.) Wiley Eastern, New Delhi, pp.219-231.

7. Gutenbrunner, C. and Jurečková, J. (1992). Regression rank scores and regression quantiles. Ann. Statist. 20, 305-330.

8. Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academia, Prague

9. Jurečková, J. (1991). Estimation in a linear model based on regression rank scores. J. Nonparametric Statist. 1, 197-203.

10. Jureckova, J. (1993). Regression rank scores, RR-tests and RR-estimates in linear regression model. Submitted for publication.

11. Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica 46, 33-50.

12. Sen, P.K. (1988a). Functional jackknifing: Rationality and general asymptotics. Ann. Statist. 16, 450-469.

13. Sen, P.K.(1988b). Functional approaches in some resampling plans: A review of some recent developments. Sankhya, Ser.A 50, 394-435.

14. Sen, P.K.(1993). Perspectives in multivariate nonparametrics: Conditional functionals and ANOCOVA models. Sankhya, Ser.A55, in press.

15. Stone, C.J. (1977). Consistent nonparametric regression.Ann.Statist. 5, 595-645.

Pranab K. Sen
Departments of Biostatistics & Statistics, University of North Carolina, Chapel Hill, NC 27599-7400.