A Relevance Weighted Nonparametric Quantile Estimator

Feifang Hu and James V Zidek

Department of Statistics

University of British Columbia

November, 1993

The research reported in this paper was supported by a grant from the Natural Science and Engineering Research Council of Canada

Abstract

This paper concerns situations in which a sample $X_1 = x_1, \cdots, X_n = x_n$ of independent observations are drawn from populations with different CDF's $F_1, \cdots, F_n$, respectively. Inference is about a quantile of another population with CDF F_0 when the data from the other populations are thought to be "relevant". Nonparametric smoothing of a quantile function would typify situations to which our theory applies. We define the relevance weighted quantile (REWQ) estimator derived from the relevance weighted empirical distribution (REWED) function. We show that the estimator has desirable asymptotic properties. A simulation study is also included. It shows that the median estimator is a robust alternative to the locally weighted averages used in conventional smoothing.

Keywords: nonparametric smoothing; quantile estimation; quantile smoothing; Bahadur representation theorem; relevance weighting.

1 Introduction

This paper concerns situations in which a sample $X_1 = x_1, \cdots, X_n = x_n$ of independent observations are drawn from populations with different CDF's $F_1, \cdots, F_n$, respectively. Inference is about an attribute θ_0 of another population with CDF F_0; an observation may be available from the latter population as well. In this paper θ_0 will be a quantile of F_0; elsewhere we address other problems of interest within the same general framework. The $\{F_i\}$ are unknown, so that we are in the nonparametric case.

The special character of the problems investigated in this problem derives from the belief that there is relevant information in the $X_i,\ i = 1, \cdots, n$ about θ_0. However this information is deemed to be "inexact". By this we mean it cannot be translated into a prior distribution from which a marginal posterior distribution for θ_0 could be constructed. And we mean there are no known structural constraints among the attributes of the various populations to force the $x_1, \cdots, x_n$ into inferences about θ_0.

The example given below illustrates the problem. That example reflects the situation underlying nonparametric regression. In fact, our approach may be thought of as generalized smoothing. In nonparametric smoothing it leads to locally weighted regression quantile estimators $\hat{\theta}_0(t), -\infty < t < \infty$ for even rough regression quantile functions $\theta_0(t), \ -\infty < t < \infty$ if the relevance of the $x(t_i)$'s corresponding to t_i's remote from t can be ascertained. Our method bears on other problems like those of meta analysis where there is no well defined underlying mathematical structure. In Section 8, we briefly discuss linkages with standard statistical methods.

We have used "information" above in a sense we believe to be consistent with Basu's use of that term (Basu 1975) when he says:

- *A problem in statistics begins with a state of nature, a parameter of interest θ about which we do not have enough information. In order to generate further information about θ, we plan and then perform a statistical experiment $\mathcal{E}$. This generates the sample x. By the term 'statistical data' we mean such a pair ($\mathcal{E}$,x) where $\mathcal{E}$ is well-defined statistical experiment and x the sample generated by a performance of the experiment. The problem of data analysis is to extract 'the whole of relevant information'–an expression made famous by R. A. Fisher–contained in the data ($\mathcal{E}$,x) about the parameter θ.*

However, statistical theory has traditionally been concerned with a narrow interpretation of the word embraced by Basu's description. [We call it "exact information" and elaborate on this concept of information in a companion paper currently in preparation (Hu and Zidek, 1993a)]. Given data, statisticians would typically construct a sampling model with a parameter θ to describe a population from which the data were supposedly drawn. Information in the sample about the population comes out through inference about θ. Alternatively, given a θ of interest the classical paradigm sees the statistician as conducting a statistical experiment to generate a sample from a population defined by a sampling distribution with parameter θ. The sample then provides information about θ. In either case, statistical inference will be based on these observations and their directly associated sampling model.

As noted above, there are problems where indirect information about θ may be used to advantage. Yet there does not seem to be a general theory underlying such problems. How to use inexact information like that encountered in metaanalysis, thus becomes an important topic which seems to have been addressed largely on a piecemeal basis.

Hu and Zidek (1993b) introduce the "relevance weighted likelihood" (REWL) as a general device for using inexact information in parameter estimation. However, when we cannot specify the population CDF's parametric form the REWL based theory is of no avail. To cope with such problems we introduce in this paper a nonparametric but general theory based on an extension of the empirical distribution function which we call the REWED ("Relevance Weighted Empirical Distribution"). We then tackle the problem of quantile estimation within that framework which is examplified by the following example.

Example 1 Distribution Smoothing. *Let $X(t_i)$ have distribution function F_{t_i}, $0 \leq t_i \leq 1$, $i = 1, \cdots, n$. Assume that $X(t_1)$, ... , $X(t_n)$ are independent and that F_t changes smoothly with t, "smooth" meaning $\sup_x |F_t(x) - F_{t+\triangle(t)}(x)| \to 0$ as $\triangle(t) \to 0$. We seek to estimate F_t for a fixed t.*

In general, let F_0 be an unknown distribution function describing the population of interest. The classical paradigm would assume independent and identically distributed (hereafter iid) observations from F_0. Here instead, only observations from other populations described by CDF's F_i, $i = 1, 2, ..., n$ are available. If we believe the $\{F_i\}$, $i = 1, 2, ..., n$ are related to F_0, $x_1, x_2,$ ... , x_n may be used for inference about attributes of F_0. The question is how.

In this paper, our answer to this question uses the REWED, defined in Section 2. From the REWED we can construct moment estimators for parameters defined in terms of the moments of F_0 [this idea will be discussed elsewhere]. But here we consider only the estimation the quantiles of F_0.

In Section 2, the REW quantile estimator will be defined in addition to REWED for the problem identified by the last example. And we will offer generalizations along with some examples.

Strong consistency of the REWED and the quantile estimators are stated in Section

3 and proved in Section 9 under mild conditions. These results generalize the results for iid sampling. some other asymptotic properties are given in Section 3.

R.R. Bahadur(1966) gives a useful asymptotic representation of the sample quantile as a simple sum of random variables by using the empirical distribution function. We give a generalization of Bahadur's results for general weights $\{p_{ni}\}$ in the non-iid case. This is the subject of Section 4.

We discuss the asymptotic normality of the REW quantile estimator in Section 5. In Section 6, we apply the theory of this paper. We get reasonable estimators for location parameters for several distributions. By comparing them with the weighted sample mean, we find their asymptotic relative efficiency (ARE) in the iid case to be fairly high. Section 7 presents the results of a simulation study, using the REW quantile estimator for the nonparametric smoothing model. The proofs of our Theorems appear in Section 9.

2 The REWED and REW Quantile Estimation

Let us reconsider Example 1. Because $t \to F_t$ changes smoothly, we could hypothetically use $\sum_{i=1}^n p_i F_{t_i}(x)$ to approximate $F_t(x)$. The choice of the weights p_i would depend on the perceived relationship between F_{t_i} and $F_t(x)$. [The $\{p_i\}$ might plausibly be generated from a kernel.] But the $\{F_{t_i}\}_1^n$ are unknown. So instead we must use the data, $X(t_i)$, $= 1, \cdots, n$ to estimate $F_t(x)$, say by $\sum_{i=1}^n p_i I(X(t_i) \leq x)$, $I(\cdot)$ being an indicator function. This empirical distribution we will call the *relevance weighted empirical distribution (REWED)*.

Estimating $F_t(x)$ by the REWED results in two errors from: (i) using $\sum_{i=1}^n p_i F_{t_i}(x)$ to approximate $F_t(x)$; (ii) using $\sum_{i=1}^n p_i I(X(t_i) \leq x)$ to estimate $\sum_{i=1}^n p_i F_{t_i}(x)$. Much of this paper will be concerned with (ii).

To generalize the ideas in the above example, let

$$\mathbf{X}_n \stackrel{def}{=} [X_{n1}, \cdots, X_{nn}];\ n \geq 1$$

be a triangular array of row-independent random variables with associated array of distribution functions, $\mathbf{F}_n \stackrel{def}{=} [F_{n1}, \cdots, F_{nn}];\ n \geq 1$ and nonnegative constants

$$\mathbf{p}_n \stackrel{def}{=} [p_{n1}, \cdots, p_{nn}];\ n \geq 1$$

satisfying $\sum p_{ni} = 1$. Define:

- the *relevance weighted empirical distribution function* (REWED) by

$$F_n(x) = \sum_{i=1}^{n} p_{ni} I(X_{ni} \leq x);$$

- the *relevance weighted average distribution function* (REWADF) for $-\infty < x < \infty$ by

$$\bar{F}_n(x) = \sum_{i=1}^{n} p_{ni} F_{ni}(x);$$

- the pth quantile of $\bar{F}_n$ by

$$\xi_{p(n)} = inf\{x : \bar{F}_n(x) \geq p\} \quad 0 < p < 1;$$

- the pth *relevance weighted quantile* (REWQ) estimator by

$$\hat{\xi}_{np} = inf\{x : F_n(x) \geq p\}$$

  for a sample $\{X_{n1}, ..., X_{nn}\}$.

To illustrate the use of these REW quantile estimators, we offer the following example.

Example 2 **Nonparametric Regression.** *Let*

$$Y_i = f(x_i) + \epsilon_i \quad x_i \in [a, b] \;\; i = 1,\ 2,\ \ldots\ ,\ n;$$

here $\epsilon_1,\ \epsilon_2,\ \ldots\ ,\ \epsilon_n$ *are iid, symmetric,* $E(\epsilon_i) = 0$ *for all* i*, and* $f(x)$ *is a smooth function. To estimate* $f(x)$ *we may use the median of the REWED. This kind of estimator is usually robust and it often quite efficient.*

3 Strong Consistency of REW Quantile Estimators

In this section, we describe strong and uniform consistency properties of the REWED. Then we describe the strong consistency of the quantile estimators derived from the REWED. In the following discussion, we assume that $F(x)$ is the CDF of interest and ξ_p its pth quantile.

Theorem 1 *(Strong Consistency of* $F_n(x)$*). a) Suppose* $\sum_{n=1}^{\infty} exp(-\epsilon^2 K_n) < \infty$ *for all* $\epsilon > 0$*, where* $K_n = (\sum_{i=1}^{n} p_{ni}^2)^{-1}$*. Then* $|F_n(x) - \bar{F}_n(x)| \to 0$ *a.s. for all* x*.*

b) Further, if $|F(x) - \bar{F}_n(x)| \to 0$ *for all* x*, then* $|F_n(x) - F(x)| \to 0$ *a.s. for all* x*.*

Corollary 1. If $log(n)/K_n = o(1)$, then

$$|F_n(x) - \bar{F}_n(x)| \to 0 \;\; a.s. \;\; for\ every\ x. \;\; \square$$

The hypothesis of the theorem is easily satisfied. If, for example, $\max_i\{p_{ni}\} = o[log(n)^{-1}]$, then the hypothesis is satisfied. The assumption $|F(x) - \bar{F}_n(x)| \to 0$ for all x is essential; without this, we cannot get a consistent estimator of the CDF $F(x)$. Qualitatively this condition is the one which gives operational meaning to the notion of "relevance weights".

Theorem 2 *(Uniformly Strong Consistency of* $F_n(x)$*) a) Under the hypothesis of Theorem 1 a), and the further assumptions that (i)* $sup_{x,n}\bar{f}_n(x)$ *is bounded and, (ii)*

$\limsup_{M\to\infty} \sup_n\{(1-\bar{F}_n(M)), \bar{F}_n(-M)\} \to 0$, *where* $\bar{f}_n(x)$ *is the derivative of* $\bar{F}_n(x)$, *then*

$$\sup_x |F_n(x) - \bar{F}_n(x)| \to 0, \quad a.s..$$

b) Further, if $\sup_x |F(x) - \bar{F}_n(x)| \to 0$, *then*

$$\sup_x |F_n(x) - F(x)| \to 0, \quad a.s..$$

When the distributions underlying our investigation derive from the same family, the conditions of the last theorem are usually satisfied.

Theorem 3 *(Strong Consistency of* $\hat{\xi}_{np}$*) Under the conditions of Theorem 2 a), suppose* $x = \xi_{p(n)}$ *solves uniquely the inequalities* $\bar{F}_n(x-) \le p \le \bar{F}_n(x)$. *Then*

$$\hat{\xi}_{np} - \xi_{p(n)} \to 0 \;\; a.s. \; for \; n \to \infty.$$

b) Further, if $\sup_x |F(x) - \bar{F}_n(x)| \to 0$, *then*

$$\hat{\xi}_{np} - \xi_p \to 0 \;\; a.s. \; for \; n \to \infty.$$

The uniqueness condition on $\xi_{p(n)}$ imposed in the last theorem cannot be dropped.

We finish this section with following theorem giving a probabilistic inequality for quantile estimators. This theorem will be used in next section.

Theorem 4 *Suppose* $x = \xi_{p(n)}$ *solves uniquely, the inequalities* $\bar{F}_n(x-) \le p \le \bar{F}_n(x)$ *for any given* $p \in (0,1)$. *Then*

$$P(|\hat{\xi}_{np} - \xi_{p(n)}| > \epsilon) \le 2exp(-2\delta_\epsilon^2(n)K_n)$$

for every $\epsilon > 0$ *and* n, *where* $\delta_\epsilon(n) = min\{\bar{F}_n(\xi_{p(n)} + \epsilon) - p, p - \bar{F}_n(\xi_{p(n)} - \epsilon)\}$.

The last theorem shows $P(|\hat{\xi}_{np} - \xi_{p(n)}| > \epsilon)$ converges to 0 exponentially fast. The value of ϵ (> 0) may depend upon K_n if desired. These bounds hold for each $n = 1,\ 2,\ ...$ and so may be applied for any fixed n as well as for asymptotic analysis.

4 Asymptotic Representation Theory.

For the case of iid data and $p_{ni} = 1/n,\ i = 1, \cdots,\ n$, R.R.Bahadur (1966) expresses sample quantiles asymptotically as sums of independent random variables by representing them as a linear transform of the sample distribution function evaluated at the relevant quantile. From these representations, a number of important properties ensue. (see Bahadur 1966 and Serfling 1980 for details). We now generalize this asymptotic representation to the cases of non-iid observations and general p_{ni}.

Theorem 5 *Let $0 < p < 0$ and $m_n = \max_{1 \le i \le n}\{p_{ni}\}$. Suppose:*

1. *$\bar{F}_n$ has bounded second derivative in the neighbourhood of $\xi_{p(n)}$ with $\bar{F}_n'(\xi_{p(n)}) = \bar{f}_n(\xi_{p(n)})$;*

2. *there exists $c > 0$, such that $\inf_n \bar{f}_n(\xi_{p(n)}) > c$;*

3. *there exists $c^* > 0$, such that $\sum_{n=1}^{\infty} K_n^{-c^*} \le \infty$;*

4. *F_{ni} has a uniformly bounded first derivative in the neighbourhood of $\xi_{p(n)}$;*

5. *$m_n = o(K_n^{-3/4}(logK_n)^{-1/4})$.*

Then

$$\hat{\xi}_{np} = \xi_{p(n)} + \frac{p - F_n(\xi_{p(n)})}{\bar{f}_n(\xi_{p(n)})} + R_n.$$

where

$$R_n = O(K_n^{-3/4}(logK_n)^{3/4}), \quad n \to \infty, \quad \textit{with probability } 1.$$

The Bahadur representation is a special case of this theorem suggesting the result of our theorem may be fairly accurate, that is hard to improve upon.

The REW sample quantile is usually hard to find, but the REW sample distribution relatively easy. By this theorem, we can use the REW sample distribution function evaluated at the relevant quantile to study the REW sample quantile asymptotically. A simple example is that we can use this representation to prove quite easily the asymptotic normality of the REW sample quantile.

5 Asymptotic Normality of $\hat{\xi}_{np}$

Except for the case of iid random variables, we cannot always find the exact distribution of $\hat{\xi}_{np}$. The asymptotic distribution of $\hat{\xi}_{np}$ given in the following theorem may therefore be useful.

Theorem 6 *Let* $0 < p < 1$ *and* $V_n = \sum_{i=1}^{n} p_{ni}^2 F_{ni}(\xi_{p(n)})(1 - F_{ni}(\xi_{p(n)}))$. *Assume* $\bar{F}_n$ *is differentiable at* $\xi_{p(n)}$, $\inf_n \bar{F}_n'(\xi_{p(n)}) > c > 0$ *and* $\max_{1 \le i \le n}(p_{ni} V_n^{-1/2}) \to 0$ *as* $n \to \infty$. *Then*

$$\lim_{n \to \infty} P(\bar{f}_n(\xi_{p(n)})(\hat{\xi}_{np} - \xi_{p(n)}) V_n^{-1/2} \le t) = \Phi(t)$$

where $\Phi(t)$ *is the distribution function of* $N(0, 1)$.

6 Applications

In this section, we use REW sample quantiles to estimate location parameters, and compare these estimators with the weighted sample mean estimators.

Example 3 *Let* $\{X_i\}$ *be an independent sample with* $X_i \sim N(\mu, \sigma_i^2)$ $i = 1, \ldots, n$, *the* σ_i^2 *being known and* μ *unknown. An estimate of* μ *is required.*

Analysis of Example 3 Using the weighted sample mean to estimate μ seems natural:

$$\hat{\mu} = \sum_{i=1}^{n} c_i X_i, \quad \sum_{i=1}^{n} c_i = 1 \text{ and } c_i \ge 0.$$

We easily deduce that $c_i = [1/\sigma_i^2]/[\sum_{j=1}^n 1/\sigma_j^2]$ minimizes the mean squared error. Then

$$\hat{\mu} \sim AN(\mu, [\sum_{i=1}^n 1/\sigma_i^2]^{-1}).$$

Now let us try using the median to estimate μ. Let F_{ni} be the distribution of X_i and $\bar{F}_n = \sum_{i=1}^n p_{ni} F_{ni}$. The median of $\bar{F}_n$ is μ and we use the sample median $\hat{\xi}_{med}$ to estimate μ. By the results of Section 5, we get

$$\hat{\xi}_{med} \sim AN(\mu, \frac{\sum_{i=1}^n p_{ni}^2}{4(\sum_{i=1}^n p_{ni} f_{ni}(\mu))^2});$$

here $f_{ni}(\mu) = (\sqrt{2\pi}\sigma_i)^{-1}$.

We want to minimize the variance of the asymptotic distribution subject to $\sum_{i=1}^n p_{ni} = 1$. We easily obtain $p_{ni} = \frac{1/\sigma_i}{\sum_{j=1}^n 1/\sigma_j}$.

The asymptotic relative efficiency of these two estimators is

$$ARE(\hat{\mu}, \hat{\xi}_{med}) = \frac{2}{\pi}$$

Remarks 1 *1. For the iid normal case, the ARE of the sample mean estimator relative to the sample median estimator is $\frac{2}{\pi}$. Here we have proved that when the samples are from normal distributions with the same mean, but different variances, the ARE of the weighted sample mean estimator relative to the weighted sample median estimator yields the same value $\frac{2}{\pi}$.*

2. The weights used in the sample mean are different from the weights used in the sample median. We only compare the two best estimators here. If we use the same weights, the ARE can be larger or smaller than $\frac{2}{\pi}$.

3. The weighted sample median should be more robust than the weighted sample mean.

Example 4 *Consider the double exponential family. Assume the density of X_i to be $1/2r_i exp(-|x-\mu|/r_i)$; the r_i are known while μ is unknown $i = 1, \cdots, n$. We again use the weighted sample mean and weighted sample median of Example 1 to estimate μ.*

Analysis of Example 4. Choose $c_i = [1/r_i^2/[\sum_{j=1}^n 1/r_j^2]$ to minimize the mean squared error. Then

$$\hat{\mu} \sim AN(\mu, 2/[\sum_{i=1}^n 1/r_i^2]).$$

By choosing $p_{ni} = \frac{1/r_i}{\sum_{j=1}^n 1/r_j}$, we get the weighted sample median $\hat{\xi}_{med}$. From the results of Section 5,

$$\hat{\xi}_{med} \sim AN(\mu, \frac{1}{\sum_{i=1}^n 1/r_i^2}).$$

The asymptotic relative efficiency of these two estimators is

$$ARE(\hat{\mu}, \hat{\xi}_{med}) = 2.$$

Remarks 2 *1. The ARE of the best sample mean estimator relative to the best sample median estimator does not depend on the* $\{r_i\}$*'s. The median is a more efficient estimator.*

2. As in the normal case, the weights used in the sample mean do not equal the weights used in the sample median.

3. The weighted sample median should be more robust.

7 Simulation Study

We have shown via asymptotics that the REW quantile estimator possesses a number of desirable asymptotic properties. In this section we use two simulated examples to obtain insight into its performance with finite sample.

The model in Example 2 is used in this simulation where we compare the REW quantile estimator with the Nadaraya-Watson estimate. The Gaussian kernel function is used to generate the relevance weights.

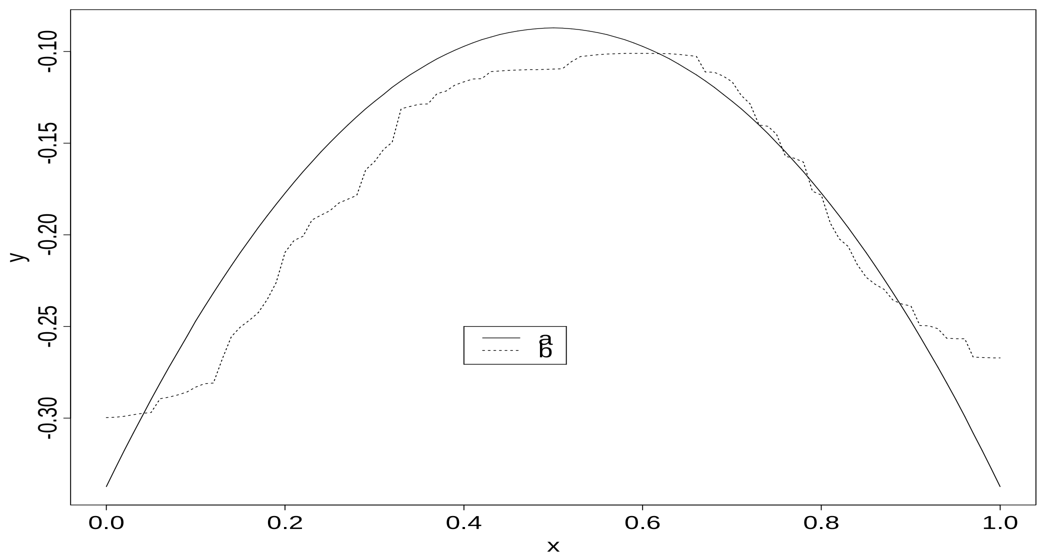

Figure 1: *A comparison of the Nadaraya-Watson estimate with REW quantile estimator. The model is* $Y = X * (1 - X) + \epsilon$*, where X is uniform (0,1) and* ϵ *is* $N(0, 0.5)$*. The sample size* $n = 1000$ *and the bandwidth,* $h = 0.1$*. The true curve is a, the REW quantile estimator b, and the Nadaraya-Watson c.*

Simulation Study 1. A random sample of size n is simulated from the model

$$Y = X(1 - X) + \epsilon,$$

with $\epsilon \sim N(0, 0.5)$ independent of $X \sim U(0, 1)$. A typical realization when $n = 1000$ is shown in Figure 1. The bandwidth used here and in all subsequances is $h = 0.1$. Let us next add 50 outliers from $N(2, 0.5)$ to the simulation experiment just described. The result is shown in Figure 2.

Simulation Study 2. In the model of Simulation 1, instead of the using the normal error, we get the ϵ from a double exponential distribution with $r = 0.1$. Figure 3 shows the results of a curve fit based on 100 simulated observations. The simulation results with 10 outliers from $N(-.5, 0.25)$ are shown in Figure 4.

For the data of Simulation 1 without outliers, the quantile curve estimate obtained by using the REW quantile estimator is shown in Figure 5.

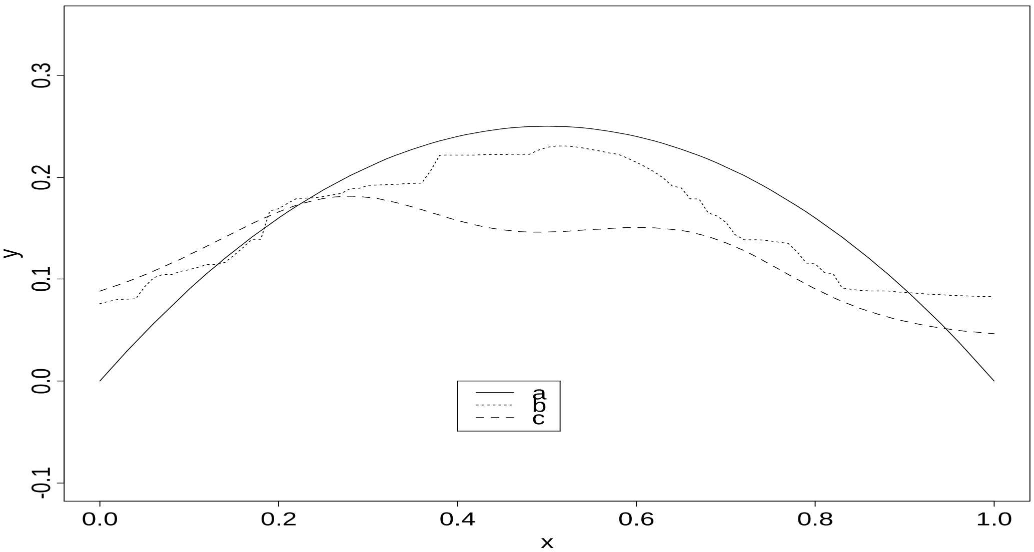

Figure 2: *A comparison of the Nadaraya-Watson estimate with REW quantile estimator with outliers. To the data depicted in Figure 1, we add* 50 ϵ*-outliers from* $N(2, 0.5)$*. The true curve is a, the REW quantile estimator b, and the Nadaraya-Watson c.*

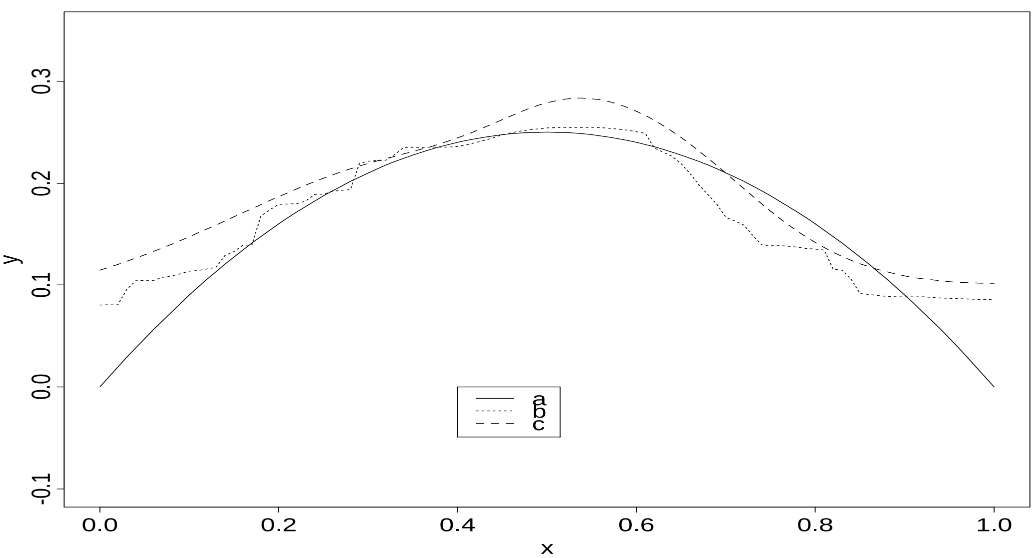

Figure 3: *A comparison of the Nadaraya-Watson estimate with REW quantile estimator. The model is* $Y = X * (1 - X) + \epsilon$*, where X is from uniform (0,1) and* ϵ *from a double exponential distribution with* $r = 0.1$*. The sample size is* $n = 100$ *and the bandwidth,* $h = 0.1$*. The true curve is a, the REW quantile estimator b, and the Nadaraya-Watson c.*

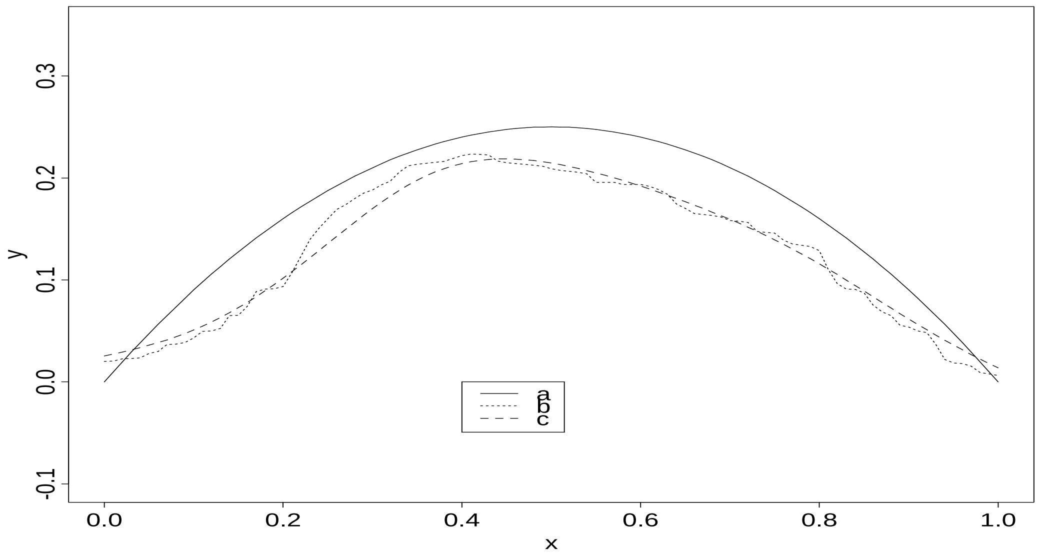

Figure 4: *A comparison of the Nadaraya-Watson estimate with REW quantile estimator with outliers. To the data depicted in Figure 3, we add* 10 *ϵ-outliers from* $N(-.5, .25)$. *The true curve is a, the REW quantile estimator b, and the Nadaraya-Watson c.*

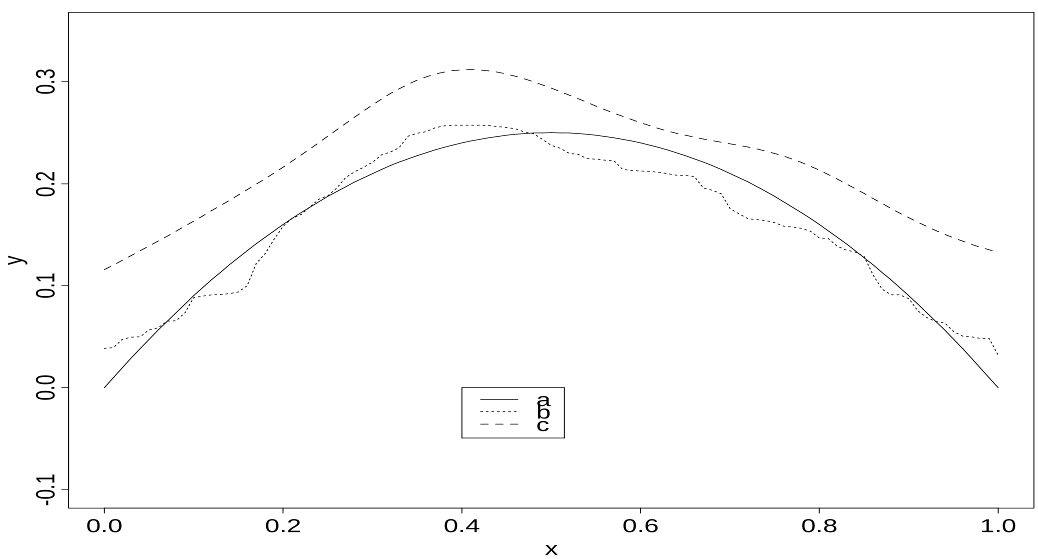

Figure 5: *A REW quantile estimator of a quantile curve. The .25 quantile curve is estimated for the data depicted in Figure 1. The true quantile curve is a, and the REW quantile estimator is b.*

The results of the simulation can be summarized as follows:

1. In the model of Example 2, when the error has the double exponential distribution, the REW quantile estimator performs a little better than Nadaraya-Watson estimate, see Figure 3. Even when the error is normal, the REW quantile estimator performs about as well as the Nadaraya-Watson estimate (see Figure 1).

2. When the data have a small fraction of outliers, say about 5 or 10 percent, the REW quantile is robust (see Figures 2 and 4). By contrast, the Nadaraya-Watson estimator fails. This observation suggests we use the REW quantile estimator and Nadaraya-Watson estimate together to diagnosis the model and determine if there are outliers in the data set. If the REW quantile estimator and Nadaraya-Watson estimate disagree, then we should reconsider the model and the outliers.

3. The REW quantile estimator seems promising judging from these simulation studies.

4. Computing the REW quantile curve estimator took about one minute in Simulation Study 1 using Splus in a Sun workstation.

8 Discussion

We have presented a general method for estimating a population quantile based on independent observations drawn from other related but not identical populations. We have shown the estimator to be strongly consistent and asymptotically normal under mild assumptions. Our method derives from a generalization of the empirical distribution (the REWED), and we have shown that the latter is also strongly consistent under certain conditions.

The context of our method includes that of nonparametric regression and smoothing. Thus our estimator may be viewed as a generalized smoothing quantile estimator. In the special case of Example 2, we obtain a nonparametric - nonparametric quantile estimator in as much as nothing is assumed about the form of the population distributions involved. In particular, as the Examples of Section 6 show, heteroscedascity is allowed in the smoothing context.

Our theory depends on the relevance weights, $\{p_{ni}\}$ used to construct the REWED. These weights express the statistician's perceived relationships among the populations and would usually be chosen on intuitive grounds. Making $\bar{F}_n$ approximate F_t well is a primary objective in this choice. Additional restrictions on the $\{p_{ni}\}$ stem from the large sample theory developed in this paper. Theorems 1, 2, and 3 on consistency, for example, require that $\sum exp(-\epsilon K_n) < \infty$ for all $\epsilon > 0$ where $K_n = (\sum_i p_{ni}^2)^{-1}$. This imposes a requirement that the $\{p_{ni}\} \rightarrow 0$ fairly rapidly as $n \rightarrow \infty$, say faster than $1/ln(n)$. And for asymptotic normality, we see in Theorem 6 the requirement that $\max_{1 \leq i \leq n}(p_{ni} V_n^{-1/2}) \rightarrow 0$ as $n \rightarrow \infty$ where $V_n = \sum_i p_{ni}^2 F_{ni}(\xi_{p(n)})(1 - F_{ni}(\xi_{p(n)})$. We believe these conditions offer some guidance on the choice of the relevance weights without unduly restricting it. In the smoothing model, we can usually use the kernel weights as the relevance weights like we did in the simulation study. The kernel weights usually satisfy the above conditions. $F_n(x)$ also arises as a population distribution estimator in finite population sampling theory (see Sarndal, Swenson and Wretman 1992, p199) where F_{ni} may be regarded as the distribution function of the subpopulation from which x_i is drawn; here $p_i = \pi_i^{-1} / \sum \pi_j^{-1}$, π_i being x_i's selection probability, $i = 1, \cdots, n$..

We would note a Bayesian connection with our theory. If the $\{p_{ni}\}$ are thought of as prior weights, then $\bar{F}_n$ is just the marginal CDF of the independent observations obtained by mixing the conditional models $\{F_{t_i}\}$. Viewed from this perspective, the weights should be chosen to make the CDF for the population of interest, F_t, that marginal CDF. We would note that incidentally this paper does provide a large sample theory for the Bayesian marginal mixture distribution, in particular for the quantiles of that mixture distribution.

9 Proofs of the Theorems

Lemma 1 *(Marcus and Zinn, 1984). Let* $\{c_n\}, n = 1, ..., \infty$, *be a sequence of real numbers and* $\{X_n\}, n = 1, ..., \infty$, *a sequence of independent random variables. Define* $U_n(t)$ *by*

$$U_n(t) = \sum_{i=1}^{n} c_i[I(X_i \le t) - P(X_i \le t)].$$

Then

$$P(|U_n(t)|(\sum_{i=1}^{n} {c_i}^2)^{-1/2} > \lambda) \le \exp(-\lambda^2/8)(1 + 2\sqrt{2\pi\lambda})$$

for all $\lambda > 0$. □

Proof of Theorem 1.

$$|F_n(x) - \bar{F}_n(x)| = |\sum_{i=1}^{n} p_{ni}[I(X_{ni} \le x) - F_{ni}(x)]| \stackrel{def}{=} |V_n(x)|,$$

say. So on applying Lemma 1 with $\lambda = \epsilon K_n^{1/2}$,

$$\begin{aligned} & P(|F_n(x) - \bar{F}_n(x)| > \epsilon) \\ = \; & P(|V_n(x)|(\sum_{i=1}^{n} p_{ni}^2)^{-1/2} > \epsilon(\sum_{i=1}^{n} p_{ni}^2)^{-1/2}) \\ \le \; & (1 + 2\sqrt{2\pi\epsilon K_n^{1/2}})exp(-\epsilon^2 K_n/8) \end{aligned}$$

for every $\epsilon > 0$. The assumption, $\sum_{n=1}^{\infty} exp(-\epsilon^2 K_n) < \infty$, implies that $K_n \to \infty$ when $n \to \infty$. It follows that for every $\epsilon > 0$, there exists N, such that for every $n > N$

$$(1 + 2\sqrt{2\pi\epsilon K_n^{1/2}}) < exp(\frac{\epsilon^2 K_n}{16}).$$

Consequently

$$(1+2\sqrt{2\pi\epsilon K_n^{1/2}})exp(-\frac{\epsilon^2 K_n}{8}) \leq exp(-\frac{\epsilon^2 K_n}{16}).$$

But $\sum_{n=1}^{\infty} exp(-\epsilon^2 K_n) < \infty$ for all $\epsilon > 0$. So $\sum_{n=1}^{\infty} exp(-\frac{\epsilon^2 K_n}{16}) < \infty$ for every $\epsilon > 0$. Hence

$$\sum_{n=1}^{\infty} P(|F_n(x) - \bar{F}_n(x)| > \epsilon) < \infty \ \ for\ all\ \epsilon > 0.$$

The Borel-Cantelli Lemma then implies

$$|F_n(x) - \bar{F}_n(x)| \to 0 \ \ a.s. \ \ for\ every\ x. \ \ \Box$$

Proof of Theorem 2. Let M be a large positive integer and

$$u_n = \max_{-M^2 \leq i \leq M^2} |F_n(i/M) - \bar{F}_n(i/M)|.$$

By Theorem 1, $u_n \to 0$ $a.s.$. Also monotonicity implies that for $(i-1)/M < t \leq i/M$

$$\begin{aligned} & F_n(t) - \bar{F}_n(t) \leq F_n[i/M] - \bar{F}_n[(i-1)/M] \\ = \ & [F_n[i/M] - \bar{F}_n[i/M]] + [\bar{F}_n[i/M] - \bar{F}_n[(i-1)/M]]. \end{aligned}$$

By similar reasoning,

$$F_n(t) - \bar{F}_n(t) \geq [F_n[(i-1)/M] - \bar{F}_n[(i-1)/M]] - [\bar{F}_n[i/M] - \bar{F}_n[(i-1)/M]].$$

So

$$\begin{aligned} & \limsup_{n\to\infty} |F_n(x) - \bar{F}_n(x)| \\ \leq \ & \limsup_{n\to\infty} u_n + \limsup_{n\to\infty} \max_{-M^2 \leq i \leq M^2} \{\bar{F}_n(\frac{i}{M}) - \bar{F}_n(\frac{i-1}{M}), 1 - \bar{F}_n(M), \bar{F}_n(-M)\} \\ \leq \ & \limsup_{n\to\infty} u_n + \limsup_{n\to\infty} M^{-1} sup_{x,n} \bar{f}_n(x) + \limsup_{n\to\infty} \{(1 - \bar{F}_n(M)), \bar{F}_n(-M)\} \end{aligned}$$

under the assumptions. Since M is arbitrary, the result follows. □

Proof of Theorem 3. Let $\epsilon > 0$. By the uniqueness condition and the definition of $\xi_{p(n)}$,

$$\bar{F}_n(\xi_{p(n)} - \epsilon) < p < \bar{F}_n(\xi_{p(n)} + \epsilon).$$

By Theorem 2, $F_n(\xi_{p(n)} - \epsilon) - \bar{F}_n(\xi_{p(n)} - \epsilon) \to 0 \;\; a.s.$ and $F_n(\xi_{p(n)} + \epsilon) - \bar{F}_n(\xi_{p(n)} + \epsilon) \to 0 \;\; a.s.$. Hence $P[F_m(\xi_{p(m)} - \epsilon) < p < F_m(\xi_{p(m)} + \epsilon),\ for\ all\ m \geq n] \to 1$ as $n \to \infty$. That is, $P(\sup_{m>n} |\hat{\xi}_{pm} - \xi_{p(m)}| > \epsilon) \to 0\ as\ n \to \infty$. This completes the proof. □

To prove the Theorem 4, we need the following useful result of Hoeffding (1963).

Lemma 2 *Let $Y_1, ..., Y_n$ be independent random variables satisfying $P(a_i \leq Y_i \leq b_i) = 1$ for each i, where $a_i < b_i$. Then for $t > 0$,*

$$P(\sum_{i=1}^{n}[Y_i - E(Y_i)] \geq t) \leq exp(-2t^2/\sum_{i=1}^{n}(b_i - a_i)^2). \quad \square$$

Proof of Theorem 4. Fix $\epsilon > 0$. Then

$$P\{|\hat{\xi}_{np} - \xi_{p(n)}| > \epsilon\} \leq P\{\hat{\xi}_{np} \geq \xi_{p(n)} + \epsilon\} + P\{\hat{\xi}_{np} \leq \xi_{p(n)} - \epsilon\}.$$

But with $Y_i = p_{ni} I(X_{ni} > \xi_{p(n)} + \epsilon)$,

$$\begin{aligned}
& P\{\hat{\xi}_{np} \geq \xi_{p(n)} + \epsilon\} \\
= \; & P\{p > F_n(\xi_{p(n)} + \epsilon)\} \\
= \; & P\{\sum_{i=1}^{n} p_{ni} I(X_{ni} > \xi_{p(n)} + \epsilon) > 1 - p\} \\
= \; & P\{\sum_{i=1}^{n}(Y_i - E(Y_i)) > 1 - p - \sum_{i=1}^{n} p_{ni}(1 - F_{ni}(\xi_{p(n)} + \epsilon))\} \\
= \; & P\{\sum_{i=1}^{n}(Y_i - E(Y_i)) > \bar{F}_n(\xi_{p(n)} + \epsilon) - p\}.
\end{aligned}$$

Because $P(0 \leq Y_i \leq p_{ni}) = 1$ for each i, by Lemma 2, we have

$$P(\hat{\xi}_{np} \geq \xi_{p(n)} + \epsilon) \leq exp(-2\delta_1^2/\sum_{i=1}^{n} p_{ni}^2) = exp(-2\delta_1^2 K_n);$$

here $\delta_1 = \bar{F}_n(\xi_{p(n)} + \epsilon) - p$. Similarly,

$$P(\hat{\xi}_{np} \leq \xi_{p(n)} - \epsilon) \leq exp(-2\delta_2^2/\sum_{i=1}^{n} p_{ni}^2) = exp(-2\delta_2^2 K_n)$$

where $\delta_2 = p - \bar{F}_n(\xi_{p(n)} - \epsilon)\}$.

Putting $\delta_\epsilon(n) = min\{\delta_1, \delta_2\}$, completes the proof. □

To prove Theorem 5, we need the following results (see Shorack and Wellner, 1986, page 855)

Lemma 3 *(Bernstein) Let $Y_1, Y_2, ..., Y_n$ be independent random variables satisfying $P(|Y_i - E(Y_i)| \leq m) = 1$, for each i, where $m < \infty$. Then, for $\epsilon > 0$,*

$$P[|\sum_{i=1}^{n}(Y_i - E(Y_i))| \geq \epsilon] \leq 2exp[-\frac{\epsilon^2}{2\sum_{i=1}^{n} Var(Y_i) + \frac{2}{3}m\epsilon}]$$

for all n = 1, 2, ...

Lemma 4 *Let $0 < p < 1$. Suppose conditions 1-3 of Theorem 5 hold. Then with probability 1 (hereafter wp1)*

$$|\hat{\xi}_{np} - \xi_{p(n)}| \leq \frac{(\sqrt{c^*/2} + 1)K_n^{-1/2}(logK_n)^{1/2}}{\bar{f}_n(\xi_{p(n)})}$$

for all sufficiently large n.

Proof. Since $\bar{F}_n$ is continuous at $\xi_{p(n)}$ with $\bar{F}_n'(\xi_{p(n)}) > 0$, $\xi_{p(n)}$ solves uniquely $\bar{F}_n(x-) \leq p \leq \bar{F}_n(x)$ and $p = \bar{F}_n(\xi_{p(n)})$. Put

$$\epsilon_n = (\sqrt{c^*/2} + 1)K_n^{-1/2}(logK_n)^{1/2}/\bar{f}_n(\xi_{p(n)}).$$

We then have

$$\begin{aligned} & \bar{F}_n(\xi_{p(n)} + \epsilon_n) - p = \bar{F}_n(\xi_{p(n)} + \epsilon_n) - \bar{F}_n(\xi_{p(n)}) \\ = & \ \bar{f}_n(\xi_{p(n)})\epsilon_n + o(\epsilon_n) \\ \geq & \ \sqrt{c^*/2}(logK_n)^{1/2}/K_n^{1/2} \end{aligned}$$

for all sufficiently large n.

Likewise we may show that $p - \bar{F}_n(\xi_{p(n)} - \epsilon_n)$ satisfies a similar inequality. Thus, with $\delta_\epsilon(n)$ as defined in Theorem 4, we have

$$2K_n\delta_\epsilon(n)^2 \geq c^* log K_n$$

for all n sufficiently large. Hence by Theorem 4,

$$P(|\hat{\xi}_{np} - \xi_{p(n)}| > \epsilon_n) \leq \frac{2}{K_n^{c^*}}$$

for all sufficiently large n.

This last result, hypothesis 3 of this theorem and the Borel-Cantelli Lemma imply that wp1 $|\hat{\xi}_{np} - \xi_{p(n)}| > \epsilon_n$ holds for only finitely many n. This completes the proof. □

Lemma 5 *Let $0 < p < 1$ and T_n be any estimator of $\xi_{p(n)}$ for which $T_n - \xi_{p(n)} \to 0$ wp1. Suppose $\bar{F}_n$ has a bounded second derivative in the neighbourhood of $\xi_{p(n)}$. Then wp1*

$$\bar{F}_n(T_n) - \bar{F}_n(\xi_{p(n)}) = \bar{F}_n'(\xi_{p(n)})(T_n - \xi_{p(n)}) + O((T_n - \xi_{p(n)})^2)$$

as $n \to \infty$.

Proof. The proof is an immediate consequence of the Taylor expansion. □

For convenience in presenting the next result, we set

$$D_n(x) = [F_n(\xi_{p(n)} + x) - F_n(\xi_{p(n)})] - [\bar{F}_n(\xi_{p(n)} + x) - \bar{F}_n(\xi_{p(n)})].$$

Lemma 6 *Let $\{a_n\}$ be a sequence of positive constants such that*

$$a_n \sim c_0 K_n^{-1/2}(log K_n)^q$$

as $n \to \infty$, for some constants $c_0 > 0$ and $q \geq 1/2$. Let $m_n = \max_{1 \leq i \leq n}\{p_{ni}\}$ and

$$H_{pn} = \sup_{|x| \leq a_n} |D_n(x)|.$$

If $m_n = o(K_n^{-3/4}(log K_n)^{(q-1)/2})$, *then under the hypothesis of Theorem 5, wp1*

$$H_{pn} = O(K_n^{-3/4}(log K_n)^{\frac{1}{2}(q+1)}).$$

Proof. Let $\{b_n\}$ be any sequence of positive integers such that $b_n \sim c_0 K_n^{1/4}(log K_n)^q$ as $n \to \infty$. For successive integers $r = -b_n, ..., b_n$, put $\eta_{r,n} = a_n b_n^{-1} r$ and $\alpha_{r,n} = \bar{F}_n(\xi_{p(n)} + \eta_{r+1,n}) - \bar{F}_n(\xi_{p(n)} + \eta_{r,n})$. The monotonicity of F_n and $\bar{F}_n$ implies that for $\eta_{r,n} \leq x \leq \eta_{r+1,n}$,

$$\begin{aligned} D_n(x) &\leq [F_n(\xi_{p(n)} + \eta_{r+1,n}) - F_n(\xi_{p(n)})] - [\bar{F}_n(\xi_{p(n)} + \eta_{r,n}) - \bar{F}_n(\xi_{p(n)})] \\ &\leq D_n(\eta_{r+1,n}) + [\bar{F}_n(\xi_{p(n)} + \eta_{r+1,n}) - \bar{F}_n(\xi_{p(n)} + \eta_{r,n})]. \end{aligned}$$

Similarly,

$$D_n(x) \geq D_n(\eta_{r,n}) - [\bar{F}_n(\xi_{p(n)} + \eta_{r+1,n}) - \bar{F}_n(\xi_{p(n)} + \eta_{r,n})].$$

So

$$H_{pn} \leq A_n + \beta_n,$$

where $A_n = \max\{|D_n(\eta_{r,n})| : -b_n \leq r \leq b_n\}$ and $\beta_n = \max\{\alpha_{r,n} : -b_n \leq r \leq b_n - 1\}$. Since $\eta_{r+1,n} - \eta_{r,n} = a_n b_n^{-1} \sim K_n^{-3/4}$, $-b_n \leq r \leq b_n - 1$, we have by the Mean Value Theorem that

$$\alpha_{r,n} \leq [\sup_{|x| \leq a_n} \bar{F}_n'(\xi_{p(n)} + x)](\eta_{r+1,n} - \eta_{r,n}) \sim [\sup_{|x| \leq a_n} \bar{F}_n'(\xi_{p(n)} + x)] K_n^{-3/4},$$

$-b_n \leq r \leq b_n - 1$. Thus

$$\beta_n = O(K_n^{-3/4}), \quad n \to \infty.$$

We now establish that wp1

$$A_n = O(K_n^{-3/4}(log K_n)^{\frac{1}{2}(q+1)}) \;\; as \; n \to \infty.$$

By the Borel-Cantelli Lemma it suffices to show that

$$\sum_{n=1}^{\infty} P(A_n \geq \gamma_n) < \infty$$

where $\gamma_n = c_1 K_n^{-3/4}(log K_n)^{\frac{1}{2}(q+1)}$ for some constant $c_1 > 0$. Now

$$P(A_n \geq \gamma_n) \leq \sum_{r=-b_n}^{b_n} P(|D_n(\eta_{r,n})| \geq \gamma_n).$$

And

$$|D_n(\eta_{r,n})| = |\sum_{i=1}^{n} p_{ni}(I(X_{ni} \in (\xi_{p(n)}, \xi_{p(n)} + \eta_{r,n})) - E(I(X_{ni} \in (\xi_{p(n)}, \xi_{p(n)} + \eta_{r,n}))))|$$

by definition. With $Y_i = p_{ni} I(X_{ni} \in (\xi_{p(n)}, \xi_{p(n)} + \eta_{r,n}))$, Bernstein's Lemma (see Lemma 3) implies

$$P(|D_n(\eta_{r,n})| \geq \gamma_n) \leq 2exp(-\gamma_n^2/D_n)$$

where $D_n = 2\sum_{i=1}^{n} Var(Y_i) + 2/3 m_n \gamma_n$.

Choose $c_2 > \sup_{n,i} f_{ni}(\xi_{p(n)})$. Then there exists an integer N such that

$$F_{ni}(\xi_{p(n)} + a_n) - F_{ni}(\xi_{p(n)}) < c_2 a_n$$

and

$$F_{ni}(\xi_{p(n)}) - F_{ni}(\xi_{p(n)} - a_n) < c_2 a_n$$

both of the above inequalities being for all $n > N$ and $i = 1, \cdots, n$. Then

$$\sum_{i=1}^{n} Var(Y_i) \leq \sum_{i=1}^{n} p_{ni}^2 c_2 a_n = K_n^{-1} c_2 a_n.$$

Hence

$$\gamma_n^2/D_n \geq \gamma_n^2/\{2K_n^{-1} c_2 a_n + 2/3 m_n \gamma_n\} \geq c_1^2 log K_n/(4c_2 c_0)$$

for all sufficiently large n. The last result obtains because of the condition $m_n = o(K_n^{-3/4}(log K_n)^{(q-1)/2})$.

Given c_0 and c_2, we may choose c_1 large enough that $c_1^2(4c_2c_0)^{-1} > c^* + 1$. It then follows that there exists N^* such that

$$P(|D_n(\eta_{r,n})| \geq \gamma_n) \leq 2K_n^{-(c^*+1)}$$

for all $|r| \le b_n$ and $n > N^*$. Consequently, for $n > N^*$

$$P(A_n \ge \gamma_n) \le 8 b_n K_n^{-(c^*+1)}.$$

In turn this implies

$$P(A_n \ge \gamma_n) \le 8 K_n^{-c^*}.$$

Hence $\sum_{n=1}^{\infty} P(A_n \ge \gamma_n) < \infty$, and the proof is complete. □

Proof of Theorem 5. Under the conditions of Theorem, we may apply Lemma 4. This means Lemma 5 becomes applicable with $T_n = \hat{\xi}_{np}$ and we have wp1,

$$\bar{F}_n(\hat{\xi}_{np}) - \bar{F}_n(\xi_{p(n)}) = \bar{f}_n(\xi_{p(n)})(\hat{\xi}_{np} - \xi_{p(n)}) + O(K_n^{-1} log K_n), \quad as\ n \to \infty.$$

Now using Lemma 6 with $q = 1/2$, and appealing to Lemma 4 again, we may pass from the last conclusion to: wp1

$$F_n(\hat{\xi}_{np}) - F_n(\xi_{p(n)}) = \bar{f}_n(\xi_{p(n)})(\hat{\xi}_{np} - \xi_{p(n)}) + O(K_n^{-3/4}(log K_n)^{3/4}), \quad as\ n \to \infty.$$

Finally, since wp1: $F_n(\hat{\xi}_{np}) = p + O(m_n)$, as $n \to \infty$, we have wp1

$$p - F_n(\xi_{p(n)}) = \bar{f}_n(\xi_{p(n)})(\hat{\xi}_{np} - \xi_{p(n)}) + O(K_n^{-3/4}(log K_n)^{3/4}), \quad as\ n \to \infty.$$

This completes the proof. □

Proof of Theorem 6. Fix t, and put

$$G_n(t) = P[\bar{f}_n(\xi_{p(n)})(\hat{\xi}_{np} - \xi_{p(n)}) V_n^{-1/2} \le t] = P[\hat{\xi}_{np} \le a_n],$$

where $a_n = \xi_{p(n)} + t V_n^{1/2} / \bar{f}_n(\xi_{p(n)})$ Then by the definition of $\hat{\xi}_{np}$

$$G_n(t) = P(F_n(a_n) \ge p).$$

Thus

$$\begin{aligned} G_n(t) &= P(\sum_{i=1}^{n} p_{ni} I(X_{ni} \le a_n) \ge p) \\ &= P(V_n^{-1/2} \sum_{i=1}^{n} p_{ni}(I(X_{ni} \le a_n) - E(I(X_{ni} \le a_n))) \ge V_n^{-1/2}[p - \sum_{i=1}^{n} p_{ni} E(I(X_{ni} \le a_n))]) \\ &= P(Z_n \ge c_n); \end{aligned}$$

here

$$Z_n = V_n^{-1/2} \sum_{i=1}^{n} p_{ni}(I(X_{ni} \le a_n) - E(I(X_{ni} \le a_n)))$$

and

$$c_n = V_n^{-1/2}.[p - \sum_{i=1}^{n} p_{ni} E(I(X_{ni} \le a_n))]$$

We first prove $Z_n \to N(0,1)$ in distribution and then that $c_n \to -t$ as $n \to \infty$ to complete the proof . To this end

$$Z_n = \sum_{i=1}^{n} p_{ni} V_n^{-1/2}(I(X_{ni} \le a_n) - F_{ni}(a_n)) = \sum_{i=1}^{n} \eta_{ni}$$

where $\eta_{ni} = p_{ni} V_n^{-1/2}(I(X_{ni} \le a_n) - F_{ni}(a_n))$.

¿From the condition $\max_{1 \le i \le n}(p_{ni} V_n^{-1/2}) \to 0$, we get $\max_{1 \le i \le n} p_{ni} \to 0$ and $V_n \to 0$. We then easily obtain for every $\epsilon > 0$ and $\tau > 0$:

1. $\sum_{i=1}^{n} P(|\eta_{ni}| \ge \epsilon) \to 0$. (since $|\eta_{ni}| \le 2 \max_{1 \le i \le n}(p_{ni} V_n^{-1/2}) \to 0$);
2. $\sum_{i=1}^{n}[E(\eta_{ni}^2 I(|\eta_{ni}| < \tau)) - (sinceE(\eta_{ni} I(|\eta_{ni}| < \tau)))^2] \to 1$. (For n large enough, $I(|\eta_{ni}| < \tau)) = 1$);
3. $\sum_{i=1}^{n} E(\eta_{ni} I(|\eta_{ni}| < \tau)) \to 0$.

So $Z_n \to N(0,1)$ in distribution by the CLT (see Chung, 1968, page 191) for triangular independent random variables.

Next we prove $c_n \to -t$.

$$\begin{aligned}
c_n &= V_n^{-1/2}[p - \sum_{i=1}^{n} p_{ni} E(I(X_{ni} \le a_n))] \\
&= V_n^{-1/2} \sum_{i=1}^{n} p_{ni}(F_{ni}(\xi_{p(n)}) - F_{ni}(a_n)) \\
&= -V_n^{-1/2} \sum_{i=1}^{n} p_{ni}(f_{ni}(\xi_{p(n)})(a_n - \xi_{p(n)}) + o(a_n - \xi_{p(n)})) \\
&\to -t \ \ as\ n \to \infty.
\end{aligned}$$

The proof is now complete. □

References

[1] Bahadur, R.R.(1966). A note on quantiles in large sample, *Ann.Math.Statist.* 37, 577-580.

[2] Chung, K.L. (1968). *A course in probability theory*, New York: Harcourt, Brace & World, Inc..

[3] Eubank, R.L.(1989). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, Inc. New York and Basel.

[4] Hardle, W. (1990). *Applied Nonparametric Regression.* New York: Cambridge University Press.

[5] Hoeffding, W.(1963). Probability Inequalities for sum of bounded random variables, *JASA* 58, 13-30.

[6] Hu, F. and Zidek, J.V.(1993a) Inexact sampling information in statistics. in preparation.

[7] Hu, F. and Zidek, J.V.(1993b) Relevance weighted likelihood and its applications. in preparation.

[8] Marcus, M.B. and Zinn, J.(1984). The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-iid case, *Ann.Prob.* 12, 335-360.

[9] Marshall, A.W. and Olkin, I.(1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

[10] Sarndal, C.-E., Swenson, B. and Wretman, J.(1992). *Model Assessed Survey Sampling*, New York: Springer-Valag.

[11] Serfling, R.J.(1980). *Approximation Theorems of Mathematical Statistics* , Wiley, New York.

[12] Serfling, R.J.(1984). Generalized L-, M- and R-statistics, *Ann.Statist.* 12, 76-86.

[13] Shorack, G.R.(1978). Convergence of reduced empirical and quantile processes with applications of order statistics in the non-iid case,*Ann.Statist.* 1, 146-152.

[14] Shorack, G.R. and Wellner J.A.(1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.