The Annals of Statistics
1992, Vol. 20, No. 4, 2168–2176

ESTIMATING CONDITIONAL QUANTILES AT THE ROOT OF A REGRESSION FUNCTION

BY HARI MUKERJEE

Wichita State University

The Robbins–Monro process $X_{n+1} = X_n - c_n Y_n$ is a standard stochastic approximation method for estimating the root θ of an unknown regression function. There is a vast literature on the convergence properties of X_n to θ. In practice, one is also interested in the conditional distribution of the system under the sequential control when the control is set at θ or near θ. This problem appears to have received no attention in the literature. We introduce an estimator using methods of nonparametric conditional quantile estimation and derive its asymptotic properties.

1. Introduction. Suppose that $F(\cdot|x)$ is a distribution function (d.f.) for each $x \in \mathbb{R}$ with finite mean $m(x)$ and that $m(x) = 0$ has a unique root θ. The Robbins–Monro (RM) (1951) procedure for sequentially estimating θ is as follows: Let X_1 be an arbitrary random variable (r.v.). Observe Y_1 at X_1, that is, Y_1 is an unbiased sample from the (random) d.f. $F(\cdot|X_1)$. For $n \geq 1$ define recursively

$$X_{n+1} = X_n - c_n Y_n, \tag{1}$$

where $\{c_n\}$ is a sequence of nonnegative constants and the conditional distribution of Y_n given $\{X_n, \dots, X_1\}$ is $F(\cdot|X_n)$.

This procedure and its many variations have been studied extensively. The convergence almost everywhere, in mean square and in probability of X_n to θ have been studied as has the asymptotic distribution of X_n; see, for example, Robbins and Monro (1951), Chung (1954), Blum (1954), Sacks (1958), Venter (1967), Fabian (1968), Goodsell and Hanson (1977), Kersting (1977) and Ruppert (1982). However, $m(\theta) = 0$ is just a one-point summary of the "performance" $F(\cdot|\theta)$ of the system under the sequential control when the control is set at the desirable value θ. In practice, one is also interested in various quantiles of $F(\cdot|x)$ for x near θ. For example, if $F(\cdot|x)$ represents the chemical content of a product at the control setting x, one might desire to obtain an average of $m(\theta)$, but it may also be of interest to know what the 95 percentile, say, of the chemical content of the product is without stopping the recursive control process. These quantiles could be approximated by the estimated quantiles of $F(\cdot|\theta)$ if $F(\cdot|x)$ varies smoothly in x near θ.

Nonparametric conditional quantile estimation using the nearest neighbor and kernel methods has been treated by Stone (1977) and Bhattacharya and

Received November 1990; revised August 1991.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62L20, 60F05, 60F15.
Key words and phrases. Conditional quantile, Bahadur representation, stochastic approximation, Robbins–Monro process, central limit theorem, law of the iterated logarithm.

Gangopadhyay (1990), among others. The basic idea is to estimate a conditional quantile at θ by the corresponding sample quantile of the "observations" at the "observation points" in a small neighborhood of θ assuming sufficient smoothness of the conditional quantile function. A Bahadur-type representation of the resulting estimator and its asymptotic normality with a norming by $n^{2/5}$ have been obtained for optimum bandwidths. However, these are based on independent identically distributed (i.i.d.) bivariate observations with the conditioning variable having a smooth positive density at θ. Nevertheless, a similar nonparametric estimator could be used for our problem. The analysis will differ from the usual nonparametric procedures for the following reasons:

1. The point θ is unknown.
2. The distribution of $\{X_n - \theta\}$ is not known nearly as precisely as in the i.i.d. case.
3. The observations $\{Y_n\}$ are not independent because of the strong dependence structure in the observation points $\{X_n\}$.

However, we enjoy the following compensating features in our problem: (i) $X_n \to \theta$ almost surely, and "fast"; (ii) the key sequence, $\{I(Y_i \le y_i) - P(Y_i \le y_i | X_i)\}$, forms a martingale difference sequence.

For a fixed $0 < p < 1$ let $\xi p \equiv \xi$ denote the p-quantile of $F(\cdot|\theta)$. Let $1 \le k_n \le n$ be a positive integer sequence and let $I_n = \{n - k_n + 1, \dots, n\}$. We write k for k_n and let $\Sigma(\cdot)_j$, without the range of summation, indicate $\Sigma_{j \in I_n}(\cdot)_j$. Define the empirical d.f. of $\{Y_i : i \in I_n\}$ by

$$\hat{F}_n(t) = \frac{1}{k} \sum I(Y_i \le t)$$

and define the estimator of ξ after n observations by

$$\begin{aligned} \hat{\xi}_n &= \text{the } [kp]\text{th order statistic of } \{Y_i : i \in I_n\} \\ &= \inf\{t : \hat{F}_n(t) \ge [kp]/k\}. \end{aligned}$$

Our main result is a Bahadur-type almost sure representation of $\hat{\xi}_n - \xi$ as an average of an i.i.d. sequence with a remainder term. The strong consistency of $\hat{\xi}_n$, its asymptotic normality and a law of the iterated logarithm follow easily from this representation.

2. Main results and proofs. Consider the following assumptions for the RM procedure (1) with $c_n = c/n$ for some $c > 0$:

C1. $m(x) = \alpha(x - \theta) + o(x - \theta)$ as $x \to \theta$, $\alpha > 0$.

C2. $m(x)(x - \theta) > 0$ for all $x \ne \theta$.

C3. $\int_{-\infty}^{\infty}[y - m(x)]^2 F(dy|x) \equiv \sigma^2(x) \to \sigma^2(\theta) \equiv \sigma^2 > 0$ as $x \to \theta$ and $EX_1^2 < \infty$.

C4. $m^2(x) + \sigma^2(x) \le K(1 + x^2)$ for some $K > 0$.

C5. $\sup_{|x-\theta|<\delta'} \int_{-\infty}^{\infty} |y - m(x)|^{2+\delta} F(dy|x) < \infty$ for some δ and $\delta' > 0$.

THEOREM 1 [Gaposhkin and Krasulina (1974)]. *If* $c\alpha > 1/2$ *then, under* C1–C5,

$$\limsup_n \frac{n^{1/2}(X_n - \theta)}{(2\log_2 n)^{1/2}} = \frac{c\sigma}{(2c\alpha - 1)^{1/2}} \quad a.s.$$

The condition $c\alpha > 1/2$ can be satisfied in practice by choosing c large enough. For estimating the p-quantile of $F(\cdot|\theta)$ we make the further assumptions:

C6. $k = o(n)$ and $1/k = o(\sqrt{\log_2 n/n})$.

C7. Assume that there exists $\varepsilon > 0$ such that (i) $f(y|x) = F_y(y|x)$ exists for $|x - \theta| < \varepsilon$; (ii) $f(\xi|\theta) > 0$, where $F(\xi|\theta) = p$, and $f_y(\xi|\theta)$ exists; and (iii) for $|x - \theta| < \varepsilon$ and $|y - \xi| < \varepsilon$, the partial derivatives $F_x(y|x)$ and $f_x(y|x)$ exist and their absolute values are uniformly bounded in y.

Note that condition C7 implies the uniqueness of ξ as the solution of $F(y|\theta) = p$.

THEOREM 2. *Under the condition* C1–C7,

$$\hat{\xi}_n - \xi = \frac{1}{kf(\xi|\theta)} \sum_{i=1}^{k} [p - I(Z_i \le \xi)] + R_n,$$

where $\{Z_i\}$ *is an i.i.d. sequence with the d.f.* $F(\cdot|\theta)$ *and*

$$R_n = O\left(\sqrt{\log_2 n/n}\right) + O(k^{-3/4}\log n) \quad a.s.$$

The estimator $\hat{\xi}_n$ of ξ is the p-quantile of the empirical d.f. $\hat{F}_n$ of $\{Y_i: i \in I_n\}$, where the conditional distribution of Y_i given the past is $F(\cdot|X_i)$. Let

$$F(\cdot) = F(\cdot|\theta), \qquad f(\cdot) = f(\cdot|\theta),$$
$$F_i(\cdot) = F(\cdot|X_i), \qquad f_i(\cdot) = f(\cdot|X_i)$$

and

$$\overline{F}_n(\cdot) = k^{-1}\sum F_i(\cdot).$$

Then it is natural to think of $\hat{\xi}_n$ as an estimator of the p-quantile ξ_n of the random d.f. $\overline{F}_n$. Note that if $|X_i - \theta| < \varepsilon$, $i \in I_n$, then

(2) $\overline{F}_n(\xi_n) = p = F(\xi)$ and ξ_n is the unique p-quantile of $\overline{F}_n$ by C7.

The discrepancy between ξ_n and ξ gives rise to a bias in addition to the random error in estimating ξ_n by $\hat{\xi}_n$. Let

$$(3)\quad \begin{gathered} b_n = c\sigma\sqrt{2\log_2 n/n(2c\sigma - 1)} \quad \text{and} \quad S_{n\rho} \doteq \{|X_n - \theta| \le \rho b_n\}, \rho > 0; \\ \text{by Theorem 1,} \quad P(S_{n\rho}^c \text{ i.o.}) = 0 \quad \text{for all } \rho > 1. \end{gathered}$$

We will frequently have occasion to prove that $P(A_n \text{ i.o.}) = 0$ for some sequence $\{A_n\}$ by showing that $P(A_n \cap S_{n2} \text{ i.o.}) = 0$ and then invoking (3), and utilizing the fact that $2b_i < \varepsilon$ for all large i so that C7 holds. To avoid repetitive arguments, we assume throughout the remainder that $|X_i - \theta| \le 2b_i < \varepsilon$ for all $i > n - k$.

We adopt the convention that all equalities (inequalities) between r.v.'s are a.s. equalities (inequalities) and all convergences of r.v.'s are a.s. convergences unless stated otherwise.

We now prove the following lemmas.

LEMMA 1. $(1/k)\Sigma\sqrt{\log_2 j/j} = O(\sqrt{\log_2 n/n})$.

PROOF. The l.h.s. $\le \sqrt{\log_2 n/(n-k)} = O(\sqrt{\log_2 n/n})$. □

LEMMA 2. *$|F(y|x) - F(y|\theta)| = O(|x-\theta|)$ uniformly in y for $|y - \xi| < \varepsilon$ as $x \to \theta$, $F(y|\theta) - F(\xi|\theta) = f(\xi|\theta)(y - \xi) + O((y-\xi)^2)$ as $y \to \xi$ and $f(\xi|x) \ge f(\xi|\theta)/2 > 0$ for all x in a neighborhood of θ.*

PROOF. Follows immediately from C7. □

LEMMA 3. *Suppose that $|y - \xi| < \varepsilon$ and $|z - \xi| < \varepsilon$. Then $\bar{F}_n(y) - F(y) = O(\sqrt{\log_2 n/n})$ uniformly in y and $\bar{F}_n(y) - \bar{F}_n(z) = O(\sqrt{\log_2 n/n}) + f(\xi)(y-z) + O((y-\xi)^2) + O((z-\xi)^2)$ uniformly in y and z.*

PROOF. By Lemmas 1 and 2, we have

$$\begin{aligned} \bar{F}_n(y) - F(y) &= \frac{1}{k}\sum[F_i(y) - F(y)] = \frac{1}{k}\sum O(|X_i - \theta|) \\ &= O\left(\sqrt{\log_2 n/n}\right) \quad \text{uniformly in } y \end{aligned}$$

and

$$\begin{aligned} \bar{F}_n(y) - \bar{F}_n(z) &= \frac{1}{k}\sum\{[F_i(y) - F(y)] + [F(y) - F(\xi)] \\ &\qquad -[F(z) - F(\xi)] - [F_i(z) - F(z)]\} \\ &= O\left(\sqrt{\log_2 n/n}\right) + f(\xi)(y-z) + O\left((y-\xi)^2\right) + O\left((z-\xi)^2\right) \end{aligned}$$

uniformly in y and z. □

LEMMA 4. *$\xi_n - \xi = O(\sqrt{\log_2 n/n})$ a.s.*

PROOF. By (2) it is sufficient to show the existence of a $D > 0$ such that

$$\bar{F}_n(\xi - Db_n) \le F(\xi) \le \bar{F}_n(\xi + Db_n) \quad \text{for all large } n.$$

Consider n large enough so that $Db_n < \varepsilon$. By Lemma 3, $|\bar{F}_n(\xi \pm Db_n) - F(\xi \pm Db_n)| \le Cb_n$ for some $C > 0$ and for all n large enough. Moreover, $F(\xi + Db_n) \ge F(\xi) + Db_n f(\xi)/2$ and $F(\xi - Db_n) \le F(\xi) - Db_n f(\xi)/2$ for all large n by C7. Thus

$$\bar{F}_n(\xi + Db_n) \ge F(\xi) + \tfrac{1}{2}Db_n f(\xi) - Cb_n$$

and

$$\bar{F}_n(\xi - Db_n) \le F(\xi) - \tfrac{1}{2}Db_n f(\xi) + Cb_n \quad \text{for all large } n.$$

Now choose $D > 2C/f(\xi)$ to complete the proof. □

The next two lemmas are generalizations to martingale differences of the well-known lemmas due to Hoeffding (1963) and Bernstein. If $U_1, U_2, \dots$ is any sequence of r.v.'s and $\{\mathscr{F}_i\}$ is an increasing sequence of sub-σ-fields such that U_i is $\mathscr{F}_i$-measurable, denote $E(U_i|\mathscr{F}_{i-1})$ by $E_i(U_i)$ and $\mathrm{Var}(U_i|\mathscr{F}_{i-1})$ by $V_i(U_i)$ for $i = 2, 3, \dots,$ and let $E_1(U_1) = E(U_1)$ and $V_1(U_1) = \mathrm{Var}(U_1)$.

LEMMA 5. *Suppose $U_1, \dots, U_n$ are Bernoulli r.v.'s. Then*

$$P\left\{\left|\sum_{i=1}^{n} [U_i - E_i(U_i)]\right| \ge nt_n\right\} \le 2\exp(-2nt_n^2).$$

PROOF. Since $E_i[U_i - E_i(U_i)] = 0$ and the range of $U_i - E_i(U_i)$ is 1 a.s. for all i, the result follows from Hoeffding's (1963) Theorem 2 and the argument given at the end of Section 2 of that paper. □

LEMMA 6. *Suppose $U_1, \dots, U_n$ are r.v.'s with $|U_i - E_i(U_i)| \le 1$ for all i. Then for any $t_n > 0$ and $v_n > 0$,*

$$P\left\{\left|\sum_{i=1}^{n} [U_i - E_i(U_i)]\right| \ge nt_n, \sum_{i=1}^{n} V_i(U_i) \le nv_n\right\} \le 2\exp(-nt_n^2/2(t_n + v_n)).$$

For n large enough the probability bounds may be taken to be:

(i) $2\exp(-nt_n^2/4v_n)$ *if* $t_n/v_n \to 0$ *and*
(ii) $2\exp(-nt_n/4)$ *if* $v_n/t_n \to 0$.

PROOF. This is just a special case of Freedman's (1975) Theorem (1.6) [see also Steiger (1969)]. The cases (i) and (ii) follow obviously from the lemma. □

LEMMA 7. *Suppose $|y - \xi| \le C\log n/\sqrt{k}$. Then for any γ, there exists M such that $n^\gamma P\{|[\hat{F}_n(y) - \hat{F}_n(\xi)] - [\bar{F}_n(y) - \bar{F}_n(\xi)]| \ge Mk^{-3/4}\log n\}$ is summable.*

PROOF. Assume that n is large enough so that $C \log n/\sqrt{k} < \varepsilon$.

Let $E_i(\cdot) = E(\cdot|X_i, \dots, X_1)$, $V_i(\cdot) = \mathrm{Var}(\cdot|X_i, \dots, X_1)$, $U_i = I(Y_i \le y) - I(Y_i \le \xi)$ and $\pi_i = F_i(y) - F_i(\xi) = E_i(U_i)$. Then

$$\left[\hat{F}_n(y) - \hat{F}_n(\xi)\right] - \left[\bar{F}_n(y) - \bar{F}_n(\xi)\right] = \frac{1}{k}\sum(U_i - \pi_i).$$

Using Lemma 3 and C6,

$$\begin{aligned}\sum V_i(U_i) &= \sum|\pi_i|(1 - |\pi_i|) \le \sum|\pi_i| = k\left|\bar{F}_n(y) - \bar{F}_n(\xi)\right| \\ &\le \sqrt{k}\, Cf(\xi)\log n + O(\log^2 n) + O\left(k\sqrt{\log_2 n/n}\right) \le \sqrt{k}\, D\log n\end{aligned}$$

for some $D > 0$ for all large n. Then Lemma 6(i) implies that for all large n,

$$\begin{aligned}&n^\gamma P\left\{\frac{1}{k}\left|\sum(U_i - \pi_i)\right| \ge Mk^{-3/4}\log n\right\} \\ &\quad = n^\gamma P\left\{\frac{1}{k}\left|\sum(U_i - \pi_i)\right| \ge Mk^{-3/4}\log n,\ \sum V_i(U_i) \le \sqrt{k}\, D\log n\right\} \\ &\quad \le 2n^\gamma \exp\left(-(M^2/4D)\log n\right),\end{aligned}$$

which is summable for M sufficiently large. □

LEMMA 8. $\hat{\xi}_n - \xi = O(\sqrt{\log n/k})$.

PROOF. Let $d_n = \sqrt{\log n/k}$. It is sufficient to show that $P\{|\hat{\xi}_n - \xi| \ge Cd_n\}$ is summable for some $C > 0$. Assume that n is large enough so that $Cd_n < \varepsilon$. Now $\hat{\xi}_n \le \xi - Cd_n$ implies that

$$\begin{aligned}&\frac{1}{k}\sum\left[I(Y_i \le \xi - Cd_n) - F_i(\xi - Cd_n)\right] \\ &\quad \ge [kp]/k - \bar{F}_n(\xi - Cd_n) \\ &\quad \ge p - \frac{1}{k} - F(\xi - Cd_n) - \left[\bar{F}_n(\xi - Cd_n) - F(\xi - Cd_n)\right] \\ &\quad \ge -\frac{1}{k} + \frac{1}{2}Cd_n f(\xi) + O\left(\sqrt{\log_2 n/n}\right) \ge \frac{1}{4}Cd_n f(\xi)\end{aligned}$$

for all large n by Lemma 3, C6 and C7. By Lemma 5,

$$\begin{aligned}&P\left\{\frac{1}{k}\left|\sum\left[I(Y_i \le \xi - Cd_n) - F_i(\xi - Cd_n)\right]\right| \ge \frac{1}{4}Cd_n f(\xi)\right\} \\ &\quad \le 2\exp\left(-kC^2 d_n^2 f^2(\xi)/8\right) = 2\exp\left(-C^2 f^2(\xi)\log n/8\right),\end{aligned}$$

which is summable for C sufficiently large. The case $\hat{\xi}_n \ge \xi + Cd_n$ is similar. □

Let

$$r_n = \log n/\sqrt{k},\ s_n = [k^{1/4}],$$

$$G_n(y) = \left[\hat{F}_n(y) - \hat{F}_n(\xi)\right] - \left[\bar{F}_n(y) - \bar{F}_n(\xi)\right]$$

and

$$H_n = \sup_{|y-\xi|\le r_n} |G_n(y)|.$$

For $-s_n \le q \le s_n$ let $\eta_{q,n} = \xi + qr_n/s_n$, and for $-s_n \le q \le s_{n-1}$ let

$$\alpha_{q,n} = \bar{F}_n(\eta_{q+1,n}) - \bar{F}_n(\eta_{q,n}) \quad \text{and} \quad J_{q,n} = [\eta_{q,n}, \eta_{q+1,n}].$$

LEMMA 9. *$H_n = R_n$, where $R_n = O(k^{-3/4}\log n) + O(\sqrt{\log_2 n/n})$.*

PROOF. From the monotonicity of $\hat{F}_n$ and $\bar{F}_n$ it follows that for $y \in J_{q,n}$,

$$G_n(\eta_{q,n}) - \alpha_{q,n} \le G_n(y) \le G_n(\eta_{q+1,n}) + \alpha_{q,n}.$$

Hence, $H_n \le K_n + \beta_n$, where

$$K_n = \max\{|G_n(\eta_{q,n})|: -s_n \le q \le s_n\}$$

and

$$\beta_n = \max\{\alpha_{q,n}: -s_n \le q \le s_{n-1}\}.$$

Assume that n is large enough so that $r_n < \varepsilon$. Then the lemma follows from the facts that $\beta_n = O(\sqrt{\log_2 n/n}) + O(r_n/s_n) + O(r_n^2) = R_n$ by Lemma 3, and $P(K_n \ge Mk^{-3/4}\log n) \le 2s_n \max\{P(|G_n(\eta_{q,n})| \ge Mk^{-3/4}\log n)\colon -s_n \le q \le s_n\}$, which is summable for M large enough by Lemma 7, implying that

$$P(K_n \ge Mk^{-3/4}\log n \text{ i.o.}) = 0 \quad \text{for } M \text{ large.} \qquad \square$$

PROOF OF THEOREM 2. From Lemmas 3, 8 and 9 and C6, we have

$$\begin{aligned} p - \hat{F}_n(\xi) &= \hat{F}_n(\hat{\xi}_n) + O(1/k) - \hat{F}_n(\xi) \\ &= \bar{F}_n(\hat{\xi}_n) - \bar{F}_n(\xi) + R_n = f(\xi)(\hat{\xi}_n - \xi) + R_n \end{aligned}$$

and thus

$$\hat{\xi}_n - \xi = \frac{1}{kf(\xi)} \sum [p - I(Y_i \le \xi)] + R_n. \tag{4}$$

For any d.f. G let $G^{-1}(t) = \inf\{x: G(x) \ge t\}$. From C7, $F_i(\cdot)$ is continuous for $|X_i - \theta| < \varepsilon$ and thus $Z_i \equiv F^{-1} \circ F_i(Y_i)$ has the conditional d.f. $F(\cdot)$ for $i \in I_n$. To show that $\{Z_i: i \in I_n\}$ is independent, we first note that

$$\begin{aligned} P(Z_i \le z_i, i \in I_n) &= E[E_n[P(Z_i \le z_i, i \in I_n)]] \\ &= F(z_n)E[E_n[P(Z_i \le z_i, i \in I_n - \{n\})]] \\ &= F(z_n)P(Z_i \le z_i, i \in I_n - \{n\}) \end{aligned}$$

and then use an induction argument.

Now $|F_i(\xi) - p| = |F_i(\xi) - F(\xi)| = O(b_{n-k}) = O(\sqrt{\log_2 n/n})$ and $f(\xi|x) \ge f(\xi)/2$ for all x in some neighborhood of θ by Lemma 2. Hence, $F_i^{-1}(p)$ is the unique solution of $F_i(y) = p$ and $Y_i \le F_i^{-1}(p) \Leftrightarrow F_i(Y_i) \le p \Leftrightarrow Z_i \le F^{-1}(p) = \xi$, $i \in I_n$, for all large n, and $|F_i^{-1}(p) - \xi| = O(\sqrt{\log_2 n/n})$. Thus, for n large enough, letting

$$U_i = I(Y_i \le \xi) - I(Z_i \le \xi) = I(Y_i \le \xi) - I[Y_i \le F_i^{-1}(p)]$$

and

$$\pi_i = E_i(U_i) = F_i(\xi) - p,$$

we have

$$\begin{aligned}\frac{1}{k}\sum[p - I(Y_i \le \xi)] &= \frac{1}{k}\sum[F_i(\xi) - I(Z_i \le \xi)] + O(k^{-3/4}\log n)\\ &= \frac{1}{k}\sum[p - I(Z_i \le \xi)] + R_n\end{aligned}$$

by Lemma 7 and the results above. Substitution in (4) of the first expression by the last in the last set of equalities completes the proof of the theorem. □

Let $\tau^2 = p(1-p)/f^2(\xi)$. The following theorem follows immediately from Theorem 2 and well-known results for averages of i.i.d. random variables.

THEOREM 3. *Under the conditions* C1–C7:

(i) $\hat{\xi}_n \to \xi$ *a.s.*;
(ii) $\limsup_n \pm (k^{1/2}/(2\log_2 n)^{1/2})(\hat{\xi}_n - \xi) = \tau$ *a.s.*; *and*
(iii) $k^{1/2}(\hat{\xi}_n - \xi) \to N(0, \tau^2)$ *in distribution, if* $k = o(n/\log_2 n)$.

3. Concluding remarks.

1. The remainder term R_n in Theorem 2 has two components—$R_{n1} = O(k^{-3/4}\log n)$, which is similar to the remainder term in Bahadur (1966) for k observations, and $R_{n2} = O(\sqrt{\log_2 n/n})$, which is due to bias and is independent of k for k as in C6. The term $R_{n2} = o(R_{n1})$ if $k = o(n^{2/3}\log^{4/3} n/\log_2^{2/3} n)$; otherwise, $R_n = R_{n2}$.
2. By C6 and Theorem 3(iii), k has to be chosen roughly between $\sqrt{n/\log_2 n}$ and $n/\log_2 n$. When $\{(X_i, Y_i)\}$ is an i.i.d. sequence with θ known, the central limit theorem for the $k-NN$ estimator uses the norming by $n^{2/5}$ for optimal k. In our case we could use a norming that is almost the same as if all the observations were at θ, but not quite. This happens because $X_n \to \theta$ at the "optimal" rate given by Theorem 1.
3. When $\{(X_i, Y_i)\}$ is i.i.d. with a smooth marginal density for X_1 near θ, Bhattacharya and Gangopadhyay (1990) show that the bias term $\xi_n - \xi$ is zero in the first order, is deterministic and proportional to $(k/n)^2$ in the second order and the remainder is $O((k/n)^3)$ a.s. using sufficient smoothness assumptions. In the RM process, X_n takes a step only of the order of $1/n$, and thus the observation points tend to stay near the same place for a long time, and we do not get cancellations of individual bias terms in the first order as in the i.i.d. case. This is related to the phenomenon giving rise to the famous arcsine law of Feller.

4. On examination of the proofs we see that the only property of the RM process we have used is the conclusion of Theorem 1. Since the quantile estimation does not interfere with the parent RM process, any modification of the RM procedure yielding the same conclusion could be used. This includes the adaptive procedures of Venter (1967), Lai and Robbins (1979) and Wei (1985).

Acknowledgment. The author wishes to express his gratitude to two referees whose comments have been very helpful.

REFERENCES

BHATTACHARYA, P. K. and GANGOPADHYAY, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

BLUM, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 382–386.

CHUNG, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25** 463–483.

FABIAN, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327–1332.

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.

GAPOSKIN, V. F. and KRASULINA, T. P. (1974). On the law of the iterated logarithm in stochastic approximation processes. *Theor. Probab. Appl.* **19** 844–850.

GOODSELL, C. A. and HANSON, D. L. (1976). Almost sure convergence for the Robbins–Monro process. *Ann. Probab.* **4** 890–901.

HOEFFDING, W. (1963). Probability inequalities for sums and bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KERSTING, G. (1977). Almost sure approximation of the Robbins–Monro process by sums of independent variables. *Ann. Probab.* **5** 954–965.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

RUPPERT, D. (1982). Almost sure approximations to the Robbins–Monro and Kiefer–Wolfowitz processes with dependent noise. *Ann. Probab.* **10** 178–187.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

STEIGER, W. (1969). A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Statist.* **40** 764–769.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

VENTER, J. H. (1967). An extension of the Robbins–Monro procedure. *Ann. Math. Statist.* **38** 181–190.

WEI, C. Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *Ann. Statist.* **13** 1498–1508.

DEPARTMENT OF MATHEMATICS AND STATISTICS
WICHITA STATE UNIVERSITY
WICHITA, KANSAS 67208-1595