The Annals of Probability
1977, Vol. 5, No. 6, 954-965

ALMOST SURE APPROXIMATION OF THE ROBBINS-MONRO PROCESS BY SUMS OF INDEPENDENT RANDOM VARIABLES[1]

BY GÖTZ KERSTING

Universität Göttingen

It is shown in this paper that the sample paths of a Robbins-Monro process with harmonic coefficients may be approximated by weighted sums of independent, identically distributed random variables. A law of iterated logarithm and a weak invariance principle follow from this result.

1. Introduction. Let H_x, $x \in \mathbb{R}$ be a family of probability measures on the real line and assume that

$$M(x) = \int y H_x(dy)$$

exists. Assume that the equation $M(x) = 0$ has a unique root θ. Robbins and Monro (1951) gave a statistical method for the estimation of θ. They chose an arbitrary random variable X_1 and defined random variables X_n recursively by

$$X_{n+1} = X_n - cn^{-1}Y_n,$$

$c > 0$, where Y_n is a random variable, the conditional distribution of which, given $X_1 = x_1, \cdots, X_n = x_n$, is H_{x_n}. Several authors proved convergence of X_n to θ under suitable conditions. Blum (1954) proved almost sure convergence of X_n to θ if X_1 has a finite second moment and the following conditions are satisfied:

CONDITION 1. There are $c, d \in \mathbb{R}^+$ such that

$$|M(x)| \leqq c + d|x| \qquad \text{for all} \quad x \in \mathbb{R}.$$

CONDITION 2. $M(x)(x - \theta) > 0$ for all $x \neq \theta$.

CONDITION 3. $\inf_{\delta_1 < |x-\theta| < \delta_2} |M(x)| > 0$ for all $0 < \delta_1 < \delta_2 < \infty$.

CONDITION 4. $\int_{-\infty}^{\infty} (y - M(x))^2 H_x(dy) \leqq \tau^2 < \infty$ for all $x \in \mathbb{R}$.

We assume these conditions throughout this paper.

The rate of convergence of $(X_n - \theta)$ is studied by several authors. An essential condition in these papers is that $\alpha = M'(\theta) > 0$. If $2c\alpha > 1$, then under suitable conditions Chung (1954), Sacks (1958), Fabian (1968) and others proved asymptotic normality of $n^{\frac{1}{2}}(X_n - \theta)$. Révész and Major proved asymptotic normality of $(n/\log n)^{\frac{1}{2}}(X_n - \theta)$ in the case $2c\alpha = 1$, and almost sure convergence

Received August 22, 1975; revised March 5, 1977.

[1] This paper was taken from my thesis.

AMS 1970 *subject classifications*. Primary 60K99, 60G17; Secondary 60F05, 60F15.

Key words and phrases. Stochastic approximation, approximation by independent random variables, invariance principle.

of $n^{c\alpha}(X_n - \theta)$ in the case $2c\alpha < 1$. Under rather restricted conditions Major (1973) was able to prove a weak form of the law of the iterated logarithm in the case $2c\alpha > 1$, which was strengthened by Révész (1974). We give a theorem about the structure of the Robbins–Monro process from which we can, incidentally, obtain results like those mentioned above.

2. Results.

We need three other conditions to formulate the main theorem.

CONDITION 5. If $x \to y$ in $\mathbb{R}$, then $H_x \to H_y$ in the weak topology.

CONDITION 6. $\alpha = M'(\theta) > 0$ and there is a $\beta > 0$ such that

$$|M(x) - \alpha(x - \theta)| = O(|x - \theta|^{1+\beta})$$

for $x \to \theta$.

For the formulation of the last condition we have to define the "inverse" F_x^{-1} of the distribution function F_x of H_x: for all $z \in (0, 1)$ define

$$F_x^{-1}(z) = \sup\{t \mid F_x(t) \leqq z\},$$

where $F_x(t) = H_x((-\infty, t))$.

CONDITION 7. There is a $\gamma \in (0, 2]$ such that

$$\int_0^1 (F_x^{-1}(z) - F_\theta^{-1}(z))^2\,dz = O(|x - \theta|^\gamma)$$

for $x \to \theta$.

Condition 6 is satisfied if $M''(\theta)$ exists, as follows from the Taylor expansion. In this case one may choose $\beta = 1$. Condition 7 is a statement on the "distance" between H_x and H_θ as $x \to \theta$.

THEOREM 1. *Let conditions 1–7 be satisfied for the family (H_x), $x \in \mathbb{R}$. Then there exists a probability space (Ω, U, P) and random variables X_n and $V_n (n = 1, 2, \cdots)$ on it such that*

(i) *X_n is a Robbins–Monro process belonging to the family (H_x). X_1 may have an arbitrary distribution with finite second moment.*

(ii) *V_n is a sequence of independent, identically distributed random variables with distribution H_θ.*

(iii) *If $c\alpha > \frac{1}{2}$ then there exists an $\varepsilon > 0$ such that*

$$n^{\frac{1}{2}}(X_{n+1} - \theta) + \frac{c}{n^{\frac{1}{2}}} \sum_{k=1}^{n} (k/n)^{c\alpha-1} V_k = O(n^{-\varepsilon})$$

almost surely.

(iv) *If $c\alpha \leqq \frac{1}{2}$, then*

$$n^{c\alpha}(X_{n+1} - \theta) + c \sum_{k=1}^{n} k^{c\alpha-1} V_k$$

is almost surely convergent.

The proof of the theorem is given in the next section. Of course, Condition 7 is the most severe of the conditions. But in many cases Condition 7 is fulfilled if $M'(\theta)$ exists. We discuss two examples:

EXAMPLE 1. Assume that H_x may be obtained from H_θ by translation, i.e.,

$$F_x(y) = F_\theta(y - M(x)) .$$

Then

$$F_x^{-1}(z) = F_\theta^{-1}(z) + M(x)$$

so that

$$\int_0^1 (F_x^{-1}(z) - F_\theta^{-1}(z))^2 \, dz = M(x)^2 .$$

Thus, Condition 7 follows from Condition 6.

EXAMPLE 2. Assume that there is an $L < \infty$ such that

$$F_x(M(x) - L) = 0 , \qquad F_x(M(x) + L) = 1 .$$

Thus

$$|F_x^{-1}(z) - M(x)| \leqq L$$

for all $z \in (0, 1)$. Because of Condition 1 we have that

$$|F_x^{-1}(z) - F_\theta^{-1}(z)| \leqq C < \infty$$

for all $z \in (0, 1)$ and all x in a neighbourhood of θ. Further assume for all $y \in \mathbb{R}$

$$F_x(y) \geqq F_\theta(y) \qquad \text{for all} \quad x < \theta$$
$$F_x(y) \leqq F_\theta(y) \qquad \text{for all} \quad x > \theta .$$

Then we get

$$F_x^{-1}(z) \leqq F_\theta^{-1}(z) \quad x < \theta ,$$
$$F_x^{-1}(z) \geqq F_\theta^{-1}(z) \quad x > \theta .$$

Thus in a neighbourhood of θ we get

$$\begin{aligned} \int_0^1 (F_x^{-1}(z) - F_\theta^{-1}(z))^2 \, dz &\leqq C \int_0^1 |F_x^{-1}(z) - F_\theta^{-1}(z)| \, dz \\ &= C \left| \int_0^1 F_x^{-1}(z) \, dz - \int_0^1 F_\theta^{-1}(z) \, dz \right| \\ &= C|M(x)| . \end{aligned}$$

Again Condition 7 follows from Condition 6.

We now give two corollaries of Theorem 1. Denote $\sigma^2 = \int_{-\infty}^{\infty} Y^2 H_\theta(dy)$.

COROLLARY 1. *Under Conditions* 1—7 (*with* $\log_2 n = \log\log\log n$, *etc.*)

(i) *if* $c\alpha > \frac{1}{2}$, *then*

$$\limsup_{n\to\infty} \frac{n^{\frac{1}{2}}(X_n - \theta)}{(\log_2 n)^{\frac{1}{2}}} = \frac{c\sigma 2^{\frac{1}{2}}}{(2c\alpha - 1)^{\frac{1}{2}}} \quad \text{a.s.},$$

(ii) *if* $c\alpha = \frac{1}{2}$, *then*

$$\limsup_{n\to\infty} \frac{n^{\frac{1}{2}}(X_n - \theta)}{(\log n \log_3 n)^{\frac{1}{2}}} = c\sigma 2^{\frac{1}{2}} \quad \text{a.s.}$$

Corollary 1 follows from Theorem 1 and a result of Chow and Teicher (1973) which gives a law of the iterated logarithm for weighted averages of independent, identically distributed random variables.

A second application is an invariance principle. Let $W(t)$ be a Brownian motion. Define the random function $Y(t)$ by

$$Y(t) = W(t^{2c\alpha-1})$$

if $c\alpha > \frac{1}{2}$.

COROLLARY 2. *Under Conditions* 1—7, *if* $c\alpha > \frac{1}{2}$ *then the random function*

$$\frac{(2c\alpha-1)^{\frac{1}{2}}}{c\sigma} t^{c\alpha} n^{\frac{1}{2}}(X_{[nt]} - \theta) ,$$

$0 \leqq t \leqq 1$, *converges weakly to* $Y(t)$, *as* $n \to \infty$.

Because of Theorem 1, one only has to prove that the random function

$$\frac{(2c\alpha-1)^{\frac{1}{2}}}{\sigma} n^{-\frac{1}{2}} \sum_{k=1}^{[nt]} (k/n)^{c\alpha-1} V_k$$

converges weakly to $Y(t)$. This may for example be done by Theorem 15.6 of Billingsley (1968) in the same manner as Donsker's theorem is proved in this book by Theorem 15.6. We omit the proof. An interesting case is the case $c\alpha = 1$. Then

$$(c\sigma)^{-1} t n^{\frac{1}{2}}(X_{[nt]} - \theta) \to W(t)$$

weakly for $0 \leqq t \leqq 1$. A consequence is an arc-sin law: If $c\alpha = 1$, then for $n \to \infty$, $0 \leqq b \leqq 1$

$$P\left(\frac{1}{n} \#\{k \mid 1 \leqq k \leqq n, X_k > 0\} < b\right) \to \frac{2}{\pi} \arcsin b^{\frac{1}{2}} .$$

3. Proof of the theorem. Without loss of generality we assume $\theta = 0$. We divide the proof into five steps.

STEP 1. In this step of the proof we construct the probability space (Ω, U, P) and the random variables X_n and V_n. First we choose a probability space (Ω_0, U_0, P_0) and on it a rv $\bar{X}_1$ such that the distribution of $\bar{X}_1$ is equal to the distribution which X_1 shall have. Further define

$$(\Omega_i, U_i, P_i) = ([0, 1], \mathscr{B}, \lambda) , \qquad i = 1, 2, \cdots ,$$

where $\mathscr{B}$ is the Borel-σ-algebra of $[0, 1]$ and λ is the Lebesgue measure on $[0, 1]$. We define

$$(\Omega, U, P) = \prod_{i=0}^{\infty} (\Omega_i, U_i, P_i) .$$

For $\omega \in \Omega$ let $\omega_n (n = 0, 1, \cdots)$ be the nth coordinate projection. Now define

$$X_1(\omega) = \bar{X}_1(\omega_0) .$$

Thus X_1 has the desired distribution. Define for all $x \in \mathbb{R}$

$$Z_x^{(n)}(\omega) = F_x^{-1}(\omega_n) , \quad n \geqq 1 ,$$
$$V_n = Z_0^{(n)} , \quad n \geqq 1 .$$

By construction the V_n are independent. Now, as is well-known, the random variable F_x^{-1} on the probability space $([0, 1], \mathscr{B}, \lambda)$ has distribution H_x. Thus V_n has a distribution H_0 and the V_n's satisfy part (ii) of the theorem. Further, from Condition 5 we get the almost sure convergence of $F_x^{-1} \to F_y^{-1}$ with respect to λ, thus

$$Z_x^{(n)} \to Z_y^{(n)} \quad \text{a.e.}$$

as $x \to y$. At last we define the σ-algebra

$$\tilde{U}_n = \prod_{i=0}^{n} U_i \subset U .$$

Now we construct X_n and Y_n by induction such that X_{n+1} and Y_n will be $\tilde{U}_n$-measurable. Suppose that we have already constructed $Y_1, \cdots, Y_{n-1}, X_1, \cdots, X_n (n \geqq 1)$ with the demanded properties of measurability. First we construct Y_n.

Define for $r \in \mathbb{N}$

$$S_r = \sum_{k=0}^{\infty} [k2^{-r}\chi_{\{k2^{-r} \leqq X_n < (k+1)2^{-r}\}} - k2^{-r}\chi_{\{-(k+1)2^r < X_n \leqq -k2^{-r}\}}]$$

where χ_A denotes the indicator function of set A. Since X_n is $\tilde{U}_{n-1}$-measurable, S_r, $r \in \mathbb{N}$, are $\tilde{U}_{n-1}$-measurable as well. Further

$$S_r \to X_n \quad \text{as} \quad r \to \infty \tag{3.1}$$

and for all $a > 0$ and all $\mu > 0$ there is a natural number N such that for all $r, r' \geqq N$

$$|S_r(\omega) - S_{r'}(\omega)| \leqq \mu \tag{3.2}$$

is true. Define the random variable

$$W_r(\omega) = Z_{S_r(\omega)}^{(n)}(\omega) .$$

We would like to show that W_r is convergent in probability. First we note, since $Z_x^{(n)} \to Z_y^{(n)}$ a.s. as $x \to y$, that for all $a > 0$ and all $\varepsilon, \delta > 0$ there is a $\mu > 0$ such that for all $x, y \in [-2a, 2a]$ with the property $|x - y| \leqq \mu$ we have

$$P(|Z_x^{(n)} - Z_y^{(n)}| \geqq \varepsilon) < \delta . \tag{3.3}$$

Now take any $\varepsilon, \delta > 0$. Choose $a > 0$ such that $P(|X_n| \geqq a) < \delta$. For ε, δ, a choose $\mu > 0$ such that (3.3) holds and for μ choose $N \in \mathbb{N}$ such that (3.2) is true. Without loss of generality we may assume $\mu \leqq a$. Then from (3.1), (3.2) we get that for all $r \geqq N$ and all ω with the property $|X_n(\omega)| \leqq a$ we have $|S_r(\omega)| \leqq 2a$.

Now choose $q, r \in \mathbb{N}$ with $N \leqq r < q$. From the construction of S_r we get that $S_r(\omega)$ is equal to a certain $t \in \mathbb{R}$ on the set $\{S_q = s\}$, $s \in \mathbb{R}$. From (3.2) it follows that $|t - s| < \mu$. Thus

$$\begin{aligned} P(|W_r - W_q| \geqq \varepsilon, S_q = s, |X_n| \leqq a) &= P(|Z_t^{(n)} - Z_s^{(n)}| \geqq \varepsilon, S_q = s, |X_n| \leqq a) \\ &= P(|Z_t^{(n)} - Z_s^{(n)}| \geqq \varepsilon) P(S_q = s, |X_n| \leqq a) \\ &\leqq \delta P(S_q = s) . \end{aligned}$$

The last equality follows from the independence between ω_n and the pair (S_r, X_n); i.e., the independence between the σ-algebras U_n and $\tilde{U}_{n-1}$. The last inequality follows from (3.3), since $s, t \in [-2a, 2a]$ and $|s - t| \leqq \mu$. Thus

$$P(|W_r - W_q| \geqq \varepsilon) \leqq P(|W_r - W_q| \geqq \varepsilon, |X_n| \leqq a) + P(|X_n| \geqq a) \leqq 2\delta .$$

Thus W_r converges in probability to some $\tilde{U}_n$-measurable function which we call Y_n. We define $X_{n+1} = X_n - cn^{-1}Y_n$. By construction Y_n and X_{n+1} are $\tilde{U}_n$-measurable, thus the induction is finished. We prove now that X_n is a Robbins–Monro process belonging to the family H_x. Let f be a continuous, bounded function on $\mathbb{R}$. We show that a.s.

$$\begin{aligned} E(f(Y_n) \mid \tilde{U}_{n-1})(\omega) &= E(f(Z^{(n)}_{X_n(\omega)})) \\ &= E(f(Z_x{}^{(n)})) \end{aligned}$$

for that value of x such that $x = X_n(\omega)$.

Take a sequence (r') of natural numbers such that $W_{r'} \to Y_n$ almost surely. From the definition of W_r we get

$$Y_n = \lim_{r' \to \infty} \sum_i Z^{(n)}_{s_i} \chi_{\{S_{r'} = s_i\}} \quad \text{a.s.}$$

Since f is continuous and bounded, we get a.s.

$$\begin{aligned} E(f(Y_n) \mid \tilde{U}_{n-1}) &= \lim_{r' \to \infty} \sum_i E(f(Z^{(n)}_{s_i}) \chi_{\{S_{r'} = s_i\}} \mid \tilde{U}_{n-1}) \\ &= \lim_{r' \to \infty} \sum_i \chi_{\{S_{r'} = s_i\}} E(f(Z^{(n)}_{s_i})) \end{aligned}$$

since $S_{r'}$ is $\tilde{U}_{n-1}$-measurable and $Z^{(n)}_{s_i}$ is U_n-measurable, thus independent of $\tilde{U}_{n-1}$. Thus there is a set $N \subset \Omega$ with $P(N) = 0$ such that for $\omega \in \Omega - N$ we have

$$\lim_{r' \to \infty} E(f(Z^{(n)}_{S_{r'}(\omega)})) = E(f(Y_n) \mid \tilde{U}_{n-1})(\omega)$$

and additionally $S_{r'}(\omega) \to X_n(\omega)$. Thus

$$\lim_{r' \to \infty} E(f(Z^{(n)}_{S_{r'}(\omega)})) = E(f(Z^{(n)}_{X_n(\omega)}))$$

because of the convergence theorem of Lebesgue and the fact that $Z_x{}^{(n)} \to Z_y{}^{(n)}$ a.s. as $x \to y$. This proves (3.4). Now by an approximation argument

$$P(Y_n \in A \mid \tilde{U}_{n-1})(\omega) = P(Z^{(n)}_{X_n(\omega)} \in A) = H_{X_n(\omega)}(A) \quad \text{a.s.}$$

Since $X_1, \cdots, X_n$ are $\tilde{U}_{n-1}$-measurable, we see that Y_n has the demanded property of a Robbins–Monro process.

STEP 2. We now look at the random variable

$$R_n = Y_n - V_n .$$

Of course we should like to prove that R_n is getting small in a certain sense. We prove

LEMMA 3.1. (i) $E(R_n \mid \tilde{U}_{n-1}) = M(X_n)$ a.s.

(ii) *There are positive numbers* p, q *such that* $E(R_n{}^2) \leqq pE(X_n{}^2) + qE(|X_n|^r)$.

PROOF. We already have proved

$$E(Y_n \mid \tilde{U}_{n-1}) = \int_{-\infty}^{\infty} y H_{X_n}(dy) = M(X_n) \quad \text{a.s.}$$

By construction V_n is independent of $\tilde{U}_{n-1}$, thus

$$E(V_n \mid \tilde{U}_{n-1}) = E(V_n) = 0 \quad \text{a.s.}$$

This proves (i).

From Condition 7 we have for $|x| \leqq \eta$

$$E((Z_x^{(n)} - Z_0^{(n)})^2) = \int_0^1 (F_x^{-1}(y) - F_0^{-1}(y))^2\,dy \leqq q|x|^\gamma .$$

For $|x| > \eta$ we get from Conditions 1 and 4

$$\begin{aligned} E((Z_x^{(n)} - Z_0^{(n)})^2) &\leqq 2E((Z_x^{(n)})^2) + 2E((Z_0^{(n)})^2) \\ &\leqq 2(\tau^2 + M(x)^2) + 2\tau^2 \\ &\leqq 4\tau^2 + 2(2c^2 + 2d^2x^2) \leqq px^2 . \end{aligned}$$

Thus we get for all $x \in \mathbb{R}$

$$E((Z_x^{(n)} - Z_0^{(n)})^2) \leqq px^2 + q|x|^\gamma .$$

Now, since $Z_0^{(n)} = V_n$, we get

$$\begin{aligned} E((W_r - V_n)^2) &= E(\textstyle\sum_i (Z_{s_i}^{(n)} - V_n)^2 \chi_{\{S_r = s_i\}}) \\ &= \textstyle\sum_i P(S_r = s_i) E((Z_{s_i}^{(n)} - V_n)^2) . \end{aligned}$$

Thus

$$E((W_r - V_n)^2) \leqq pE(S_r^2) + qE(|S_r|^\gamma) . \tag{3.5}$$

We used the independence of $Z_s^{(n)}$ and S_r. Now $|S_r| \leqq |X_n|$ for all $r \in \mathbb{N}$. Since $E(X_n^2) < \infty$ and $\gamma \leqq 2$, we get from the convergence theorem of Lebesgue

$$\lim_{r\to\infty} E(S_r^2) = E(X_n^2) \tag{3.6}$$

and

$$\lim_{r\to\infty} E(|S_r|^\gamma) = E(|X_n|^\gamma) . \tag{3.7}$$

On the other hand, since $W_{r'} \to Y_n$ a.s., we get by Fatou's lemma

$$E(R_n^2) \leqq \liminf_{r'\to\infty} E((W_{r'} - V_n)^2) . \tag{3.8}$$

From (3.5)–(3.8) (ii) follows.

STEP 3. Now we are able to prove statement (iii) of the theorem if we assume the following additional property

$$K_1|x| \leqq |M(x)| \leqq K_2|x| \qquad \text{for all} \quad x \in \mathbb{R} , \tag{3.9}$$

with $0 < K_1 < K_2 < \infty$. We shall consider the slightly more general situation where

$$X_{n+1} = X_n - c(n + N)^{-1} Y_n \tag{3.10}$$

with $N \geqq 0$, $c > 0$. Of course, this modification has no effect with respect to our construction in Step 1. Now as shown by Chung (1954, page 475) and Venter (1966, page 1544), we get the following estimate for $E(X_n^2)$:

LEMMA 3.2. *For a Robbins–Monro process, given by the iteration* (3.10), *which satisfies the property* (3.9) *and Condition* 4

$$E(X_n^2) = O(n^{-1}) \qquad \text{if} \quad c > 1/(2K_1)\,,$$
$$E(X_n^2) = O(n^{-2K_1 c}) \qquad \text{if} \quad c < 1/(2K_1)\,.$$

In the following we shall assume

$$c > 1/(2K_1)\,. \tag{3.11}$$

Put $A = c\alpha$. Then we get from (3.10)

$$\begin{aligned}(k+N)^A X_{k+1} &= (k+N)^A X_k - c(k+N)^{A-1} Y_k \\ &= (k+N-1)^A X_k + A(k+N)^{A-1} X_k \\ &\quad + c_k(k+N)^{A-2} X_k - c(k+N)^{A-1} Y_k\,,\end{aligned}$$

where c_k is a bounded sequence. Summing up these equations from $k = 1, \cdots, n$, we get

$$\begin{aligned}(n+N)^A X_{n+1} = N^A X_1 &+ \sum_{k=1}^n A(k+N)^{A-1} X_k \\ &+ \sum_{k=1}^n c_k(k+N)^{A-2} X_k - c\sum_{k=1}^n (k+N)^{A-1} Y_k\,.\end{aligned}$$

From $Y_k = V_k + R_k$ and from Lemma 3.1 (i) it follows that

$$\begin{aligned}(n+N)^A X_{n+1} &+ c\sum_{k=1}^n (k+N)^{A-1} V_k \\ &= N^A X_1 + \sum_{k=1}^n (k+N)^{A-1}(AX_k - cM(X_k)) \\ &\quad + \sum_{k=1}^n c_k(k+N)^{A-2} X_k \\ &\quad + c\sum_{k=1}^n (k+N)^{A-1}(E(R_k \mid \tilde{U}_{k-1}) - R_k)\,.\end{aligned} \tag{3.12}$$

We estimate the magnitude of the sums on the right side of this equation:

a) First look at $\sum_{k=1}^n (k+N)^{A-1}|X_k|^{1+\beta}$, where β is choosen according to Condition 6. Without loss of generality we may assume $\beta \leqq 1$. Then from Lemma 3.2 and the fact that $(E|X|^t)^{1/t}$ is a nondecreasing function of t we get

$$E(|X_k|^{1+\beta}) = O(k^{-\frac{1}{2}-\beta/2})\,.$$

Choose $\varepsilon_1 > 0$. Then we get

$$\frac{(k+N)^{A-1}E(|X_k|^{1+\beta})}{k^{A+\varepsilon_1}k^{-\frac{1}{2}-\beta/2}} \leqq \frac{C}{k^{1+\varepsilon_1}}\,,$$

thus

$$\sum_{k=1}^\infty \frac{(k+N)^{A-1}|X_k|^{1+\beta}}{k^{A-\frac{1}{2}-\beta/2+\varepsilon_1}} < \infty \quad \text{a.s.}$$

Now $K_1 \leqq M'(0) = \alpha$, thus $K_1 c \leqq A$, thus from (3.11) $A - \frac{1}{2} > 0$. Therefore we may choose β so small that $\beta > 0$, $A - \frac{1}{2} - \beta/2 > 0$. (Of course Condition 6 remains true!). Thus by Kronecker's lemma

$$n^{-A+\frac{1}{2}+\beta/2-\varepsilon_1}\sum_{k=1}^n (k+N)^{A-1}|X_k|^{1+\beta} \to 0$$

almost surely. From Conditions 1–4 we have by Blum's theorem that $X_n \to 0$

a.s. Thus from Condition 6 we get

$$n^{-A+\frac{1}{2}+\beta/2-\varepsilon_1} \sum_{k=1}^{n} (k+N)^{A-1}|AX_k - cM(X_k)| \to 0$$

almost surely since $A = cM'(0)$. Now choose ε_1 such that $\beta/2 - \varepsilon_1 = \varepsilon_2 > 0$. Thus we get

$$|\sum_{k=1}^{n} (k+N)^{A-1}(AX_k - cM(X_k)| = O(n^{A-\frac{1}{2}-\varepsilon_2}) .$$

Similarly

$$|\sum_{k=1}^{n} c_k(k+N)^{A-2}X_k| = O(n^{A-\frac{1}{2}-\varepsilon_3})$$

with $0 < \varepsilon_3 < \frac{1}{2}$.

b) From Lemma 3.1 (ii) and Lemma 3.2 follows as before

$$E(R_k^2) = O(k^{-\gamma/2})$$

since $\gamma \leqq 2$, thus

$$\frac{(k+N)^{2A-2}E(R_k^2)}{k^{2A-1+\varepsilon_4-\gamma/2}} \leqq Dk^{-1-\varepsilon_4}$$

with $\varepsilon_4 > 0$. By means of a well-known convergence theorem (Loève (1955) page 387) this implies that

$$\sum_{k=1}^{\infty} \frac{(k+N)^{A-1}}{k^{A-\frac{1}{2}+\varepsilon_4/2-\gamma/4}} (R_k - E(R_k \mid \tilde{U}_{k-1}))$$

is almost surely convergent (R_k is $\tilde{U}_k$-measurable!). Again, without violating Condition 7, we may choose γ so small that $A - \frac{1}{2} > \gamma/4 > 0$. By means of Kronecker's lemma we get

$$n^{-A+\frac{1}{2}+\gamma/4-\varepsilon_4/2} \sum_{k=1}^{n} (k+N)^{A-1}(R_k - E(R_k \mid \tilde{U}_{k-1}))$$

is almost surely convergent to 0. Take ε_4 so small that $\gamma/4 - \varepsilon_4/2 = \varepsilon_5 > 0$. Then

$$\sum_{k=1}^{n} (k+N)^{A-1}(R_k - E(R_k \mid \tilde{U}_{k-1})) = O(n^{A-\frac{1}{2}-\varepsilon_5}) .$$

From the estimates of a) and b) and from (3.12) we obtain

$$(3.13) \qquad (n+N)^{\frac{1}{2}}X_{n+1} + c(n+N)^{-\frac{1}{2}} \sum_{k=1}^{n} \left(\frac{k+N}{n+N}\right)^{A-1} V_k = O(n^{-\varepsilon}) \quad \text{a.s.},$$

with $\varepsilon = \min(\varepsilon_2, \varepsilon_3, \varepsilon_5) > 0$. Note that ε depends only on A, γ, β. This will be of use in the next step.

STEP 4. We now prove statement (iii) of the theorem. Assume $c > 1/(2\alpha)$. Since $M'(0)$ exists, there is a $\delta > 0$ such that for $|x| \leqq \delta$ we have

$$K_1|x| \leqq |M(x)| \leqq K_2|x| .$$

Since $c\alpha > \frac{1}{2}$, one may choose $K_1 c > \frac{1}{2}$ if δ is chosen small enough; then $c > 1/(2K_1)$. Thus K_1 fulfills (3.11). We now introduce a new family $\hat{H}_x$ of probability measures on $\mathbb{R}(x \in \mathbb{R})$ by the definition:

$$\hat{H}_x = H_x \quad \text{for} \quad |x| \leqq \delta ,$$
$$\hat{H}_x(A) = H_\delta(A - (x - \delta)K_1) \quad \text{for} \quad x > \delta ,$$
$$\hat{H}_x(A) = H_{-\delta}(A - (x + \delta)K_1) \quad \text{for} \quad x < -\delta ,$$

for all Borel-sets $A \subset \mathbb{R}$. Define $\hat{M}(x) = \int y\hat{H}_x(dy)$, thus

$$K_1|x| \leqq |\hat{M}(x)| \leqq K_2|x|$$

for all $x \in \mathbb{R}$. It is easy to see that $\hat{H}_x$ satisfies Conditions 1–7; (3.9) is satisfied, too.

Now construct according to Step 1 the Robbins–Monro process. From Conditions 1–4 we get $X_n \to 0$ almost surely. Thus for any $\delta' > 0$ there is a number $N \geqq 0$ such that $P(\sup_{n \geqq N} |X_n| \geqq \delta) \leqq \delta'$. Now define

$$(\hat{\Omega}_0, \hat{U}_0, \hat{P}_0) = \prod_{i=0}^{N} (\Omega_i, U_i, P_i) ,$$
$$(\hat{\Omega}_i, \hat{U}_i, \hat{P}_i) = (\Omega_{i+N}, U_{i+N}, P_{i+N}) , \quad i \geqq 1 ,$$
$$\hat{X}_1 = X_{N+1} ,$$
$$\hat{V}_i = V_{N+i} , \quad i \geqq 1 .$$

On this new probability space we now construct, as in Step 1, random variables $\hat{X}_n$, $\hat{Y}_n$ such that

$$\hat{X}_{n+1} = \hat{X}_n - c(n + N)^{-1}\hat{Y}_n$$

and such that $\hat{X}_n$ becomes a Robbins–Monro process with respect to $\hat{H}_x$. From the construction of $\hat{H}_x$ we get

$$\hat{Z}_x^{(i)} = Z_x^{(N+i)} \quad \text{for} \quad |x| \leqq \delta .$$

An induction argument shows that $\hat{Y}_k = Y_{k+N}$ and $\hat{X}_k = X_{k+N}$ for all $k \geqq 1$ on the set $M_n = \{\sup_{n \geqq N} |X_n| \leqq \delta\}$. Now by (3.13) we have for $\hat{X}_n$

$$(n + N)^{\frac{1}{2}}\hat{X}_{n+1} + c(n + N)^{-\frac{1}{2}} \sum_{k=1}^{n} \left(\frac{k + N}{n + N}\right)^{A-1} \hat{V}_k = O(n^{-\varepsilon}) .$$

Thus

$$(n + N)^{\frac{1}{2}}X_{n+N+1} + c(n + N)^{-\frac{1}{2}} \sum_{k=1}^{n} \left(\frac{k + N}{n + N}\right)^{\frac{1}{2}} V_{k+N} = O(n^{-\varepsilon})$$

a.s. on the set M_N. At last

$$(n + N)^{-\frac{1}{2}} \sum_{k=1}^{N} (k/n)^{A-1}V_k = O(n^{-\varepsilon'}) \quad \text{a.s.}$$

with $\varepsilon' = A - \frac{1}{2} > 0$. Thus

$$n^{\frac{1}{2}}X_{n+1} + cn^{-\frac{1}{2}} \sum_{k=1}^{n} (k/n)^{A-1}V_k = O(n^{-\varepsilon''})$$

a.s. on M_N. Now ε, ε' are dependent only from A, β, γ. Thus ε'' is independent of N. Since $P(M_N) \geqq 1 - \delta'$, the approximation holds almost surely. Thus (iii) is proved.

STEP 5. We now prove part (iv) of the theorem. The proof is similar to the proof of (iii). Again we may assume (3.9) and additionally we may assume that

$$K_1(1+\beta) > M'(0) = \alpha > K_1$$

(see Step 4). Now we get $c < 1/(2K_1)$, since $c \leqq 1/2\alpha$. From Lemma 3.2 we get for $\beta \leqq 1$

$$E(|X_k|^{1+\beta}) = O(k^{-(1+\beta)cK_1}) .$$

Since $(1+\beta)K_1 c > \alpha c = A$, we get that

$$\sum_{k=1}^{\infty} (k+N)^{A-1}|X_k|^{1+\beta} < \infty \quad \text{a.s.}$$

From Condition 6 we get that

$$\sum_{k=1}^{\infty} (k+N)^{A-1}(AX_k - cM(X_k))$$

is almost surely convergent. Thus the first sum in (3.12) is convergent. An easy argument shows that the second sum is convergent. To get the convergence of the third sum, we only have to show that

$$\sum_{k=1}^{\infty} (k+M)^{2A-2}E(R_k^2) < \infty$$

because of the convergence theorem used above. Since $A \leqq \frac{1}{2}$ we have $2A - 2 \leqq -1$. By means of Lemma 3.1 (ii) and Lemma 3.2 we get the desired result. Thus the proof is complete.

REMARK. Condition 5 was only necessary for the approximation of X_n by S_r and of Y_n by W_r. If all X_n are discrete then one may omit this condition. This occurs for example if X_1 is discrete and H_x is concentrated on a denumerable set for all $x \in \mathbb{R}$.

Acknowledgement. I wish to thank Professor Ulrich Krengel for his constant help and interest during the development of this paper.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

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

CHOW, Y. S. and TEICHER, H. (1973). Iterated logarithm laws for weighted averages. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **26** 87–94.

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

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

LOÈVE, M. (1955). *Probability Theory.* Van Nostrand, New York.

MAJOR, P. (1973). A law of iterated logarithm for the Robbins-Monro method. *Studia Sci. Math. Hungar.* **8** 92–102.

MAJOR, P. and RÉVÉSZ, P (1973). A limit theorem for the Robbins-Monro approximation. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **27** 79–86.

RÉVÉSZ, P. (1974). On the rate of convergence of Kesten's "Accelerated Stochastic Approximation." Preprint of the Hungarian Academy of Sciences, Budapest.

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

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

VENTER, J. H. (1966). On Dvoretzky's stochastic approximation theorems. *Ann. Math. Statist.* **37** 1534–1544.

INSTITUT FÜR MATHEMATISCHE STATISTIK
UNIVERSITY OF GÖTTINGEN
LOTZESTRASSE 13
D-3400 GÖTTINGEN
FEDERAL REPUBLIC OF GERMANY