ON ROBBINS–MONRO STOCHASTIC APPROXIMATION
METHOD WITH TIME VARYING OBSERVATIONS

Watanabe, Masafumi
Department of Applied Mathematics, Fukuoka University

https://doi.org/10.5109/13096
ON ROBBINS-MONRO STOCHASTIC APPROXIMATION
METHOD WITH TIME VARYING OBSERVATIONS

By

Masafumi WATANABE*

(Received October 7, 1974)

§ 1. Introduction and Summary.

The method of stochastic approximation proposed by Robbins and Monro [8] is described by the following situations. Suppose that a random variable \( Y(x) \) can be observed at \( x \) and the expected value of \( Y(x) \), denoted by \( E[Y(x)] \) (regression function), is unknown to us. Assuming that the equation \( E[Y(x)] = 0 \) has the unique root \( x = \theta \), it is desired to estimate \( \theta \) on the basis of observed values \( \{ Y(x_i) \} \) at the points \( x_1, x_2, \ldots \) which are produced by the following recurrence relation

\[
x_{n+1} = x_n - \alpha_n Y_n,
\]

where \( y_n \) is a random variable whose conditional distribution given \( (x_1, x_2, \ldots, x_n) \) coincides with the distribution of random variable \( Y(x_n) \) and \( \{ \alpha_n \} \) is a sequence of positive numbers which converges to zero as \( n \to \infty \).

In this paper, we shall assume that the random variable \( Y(x) \) depends on time \( n \), denoted by \( Y_n(x) \), and we shall consider two problems. The first is to find the root of the equation \( E[Y_n(x)] = 0 \) for sufficient large \( n \). Regarding this problem, V. Dupac [3] and T.V. Young and R.A. Westerberg [12] assumed that the equation \( E[Y_n(x)] \) has the unique root. In this paper, however, we do not assume the equation has the unique root.

The second problem is as follows. Let \( M(x) \) be an unknown function. Suppose that in place of a random variable \( Y(x) \) satisfying \( E[Y(x)] = M(x) \), we can observe, at each time \( n \), a random variable \( Y_n(x) \) for which the equation \( E[Y_n(x)] = M(x) \) may not hold true. Under these situations, then, we shall try to find the root of the equation \( M(x) = 0 \), denoted by \( x = \theta \), supposing that the root may not be unique. The second problem cannot be reduced to the first one, because in the second the equation \( E[Y_n(x)] = 0 \) may not have any root.

In this paper we shall treat the case when the observations are taken from the space of real numbers, and the case when from a Hilbert space. The problems of finding the solution of some kinds of functional equations can be reduced to our problem in the case when the observations are taken from a Hilbert space. For example,
the density estimation problems ([7], [11]) and the learning problems for a pattern classification ([9], [13], [14], [15]) can be regarded as the application of the Robbins-Monro stochastic approximation procedure in a Hilbert space.

This paper consists of six sections. In Section 2, we shall give some preliminaries and lemmas to be used throughout the paper. In Section 3, we shall treat the Robbins-Monro (abbreviated as RM) procedure in the case when the observations are taken from one-dimensional Euclidean space \( R \), and in Section 4, the case when the observations taken from a real separable Hilbert space \( H \). In Section 5, we shall give some applications of Section 4.

§ 2. Preliminaries.

Throughout this paper \( R \) denotes the one-dimensional Euclidean space, \( R^n \) the \( n \)-dimensional Euclidean space and \( H \) a real separable Hilbert space with an inner product \( (\cdot, \cdot) \) and a norm \( \|\cdot\| \). And \( H^n \) denotes the cartesian product of \( n \) spaces \( H \).

Let \( \mathcal{F} \) be the \( \sigma \)-field of subsets of \( H \) generated by the open sets. Let \( (\Omega, \mathcal{F}, P) \) be a probability space. A random element \( X \) in \( H \) is a measurable mapping from \( (\Omega, \mathcal{F}) \) into \( (H, \mathcal{F}) \).

If \( X, Y \) are random elements in \( H \) and \( h \) an element of \( H \), then \( \|X\|, (X, Y), (h, X) \) are real-valued random variables. Denoting the expectation operator by \( E \), if \( E \|X\| < \infty \), then \( EX \) is defined by the requirement \( E(h, X) = (h, EX) \) for all \( h \) in \( H \). Similarly, let \( \mathcal{B} \) be a sub-\( \sigma \)-field of \( \mathcal{F} \), then the conditional expectation of \( X \) given \( \mathcal{B} \), denoted by \( E[X|\mathcal{B}] \), is defined by the requirement \( E[(h, X)|\mathcal{B}] = (h, E[X|\mathcal{B}]) \) a.s. for all \( h \) in \( H \). The expectation operator and the conditional expectation operator defined above have the usual properties.

The following lemmas will be needed throughout this paper.

**Lemma 1** (Dvoretzky [5]). Let \( N \) be a positive integer and \( T_n, n = 1, 2, \ldots, \) be measurable Transformations of \( R^n \) into \( R \). Let \( \{\alpha_n\}, \{\beta_n\}, \{\gamma_n\} \) be sequences of non-negative numbers and \( \gamma_n \) be a non-negative valued measurable function on \( R^n \). Suppose that the following conditions are satisfied:

\[(2.1) T_n(r_1, \cdots, r_n) \leq \max \{\alpha_n, (1+\beta)\gamma_n(r_1, \cdots, r_n)+\gamma_n\}
\]

for all \( (r_1, \cdots, r_n) \in R^n \) and all \( n \geq N \),

\[(2.2) \lim_{n \to \infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \beta_n < +\infty, \quad \sum_{n=1}^{\infty} \gamma_n < +\infty
\]

and

\[(2.3) \sum_{n=1}^{\infty} \gamma_n(r_1, \cdots, r_n) = +\infty
\]

uniformly for all sequences \( \{r_n\} \subseteq R \) for which \( \sup_n r_n < \infty \). And let \( X_i \) be any constant and \( \{U_n\} \) be a sequence of random variables. Define, for each \( n \geq 1 \),

\[(2.4) X_{n+1} = T_n(x_1, \cdots, x_n) + U_n
\]

Then the conditions, \( E|X_N|^2 < +\infty \),
On Robbins-Monro Stochastic Approximation Method

\[ \sum_{n=1}^{\infty} E |U_n|^2 < +\infty \]

and

\[ E[U_n | X_1, \ldots, X_n] = 0 \quad a.s. \quad for \ all \ n \geq N \]

imply

\[ \lim_{n \to \infty} E |X_n|^2 = 0 \]

and

\[ \lim_{n \to \infty} X_n = 0 \quad a.s. \]

The following lemma is a direct application of [10].

**Lemma 2 (Venter [10]).** Let \( N \) be a positive integer and for each \( n \), \( T_n \) be a measurable Transformation of \( H^n \) into \( H \). Let \( \{\alpha_n\}, \{\beta_n\}, \{\delta_n\} \) be sequences of non-negative numbers and \( \gamma_n \) be a non-negative valued measurable function on \( H^n \). Suppose that the following conditions are satisfied:

\[ \|T_n(h_1, \ldots, h_n)\| \leq \max \{\alpha_n, (1+\beta_n)\|h_n\|^2 - \gamma_n(h_1, \ldots, h_n) + \delta_n\} \]

for all \( (h_1, \ldots, h_n) \in H^n \) and all \( n \geq N \),

\[ \lim_{n \to \infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \beta_n < +\infty, \quad \sum_{n=1}^{\infty} \delta_n < +\infty \]

and

\[ \sum_{n=1}^{\infty} \gamma_n(h_1, \ldots, h_n) = +\infty \]

uniformly for all sequences \( \{h_n\} \subset H \) for which \( \sup_n \|h_n\| < +\infty \). Let \( X_1 \) be an arbitrary element in \( H \) and \( \{U_n\} \) be a sequence of random elements in \( H \). Define, for each \( n \geq 1 \),

\[ X_{n+1} = T_n(X_1, \ldots, X_n) + U_n. \]

Then the conditions, \( E\|X_N\|^4 < \infty \),

\[ \sum_{n=1}^{m} E\|U_n\|^4 < +\infty \]

and

\[ E[U_n | X_1, \ldots, X_n] = 0 \quad a.s. \quad for \ all \ n \geq N \]

imply

\[ \lim_{n \to \infty} E \|X_n\|^2 = 0 \]

and

\[ \lim_{n \to \infty} \|X_n\| = 0 \quad a.s. \]

**Lemma 3.** Let \( \{X_n\} \) be a sequence of random variables on a probability space \( (\Omega, \mathcal{A}, P) \). Let \( \{\mathcal{A}_n\} \) be a sequence of \( \sigma \)-fields, \( \mathcal{A}_n \subset \mathcal{A}_{n+1} \subset \mathcal{A} \), where \( X_n \) is measurable with respect \( \mathcal{A}_n \) for each \( n \geq 1 \). And let \( N \) be a positive integer and \( \{\alpha_n\}, \{\nu_n\} \) be two sequences of positive numbers which satisfy there exists a number \( 0 < \lambda < 1 \) such that
\[(2.17) \quad (1-a_{n+1})\left(\frac{v_n}{v_{n+1}}\right)^{\lambda} \leq 1 \quad \text{for all } n \geq N\]

and
\[(2.18) \quad \sum_{n=1}^{\infty} v_n^{\lambda-1} < +\infty .\]

Suppose that the following conditions are satisfied
\[(2.19) \quad 0 \leq X_n \quad \text{a. s. for all } n \geq 1\]
\[(2.20) \quad E[X_{n+1} | A_n] \leq (1-a_{n+1})X_n + M_1 \cdot V_{n+1} \quad \text{a. s. for all } n \geq 1\]

where \(M_1\) is a positive constant and \(\{V_n\}\) is a sequence of random variables.

(1) If the following conditions hold
\[(2.21) \quad E[X_N] < \infty\]
and
\[(2.22) \quad E|V_n| \leq v_n \quad \text{for all } n \geq N\]

then there exists a constant \(C > 0\) such that
\[(2.23) \quad E[X_n] \leq C \cdot v_n^{\lambda} \quad \text{for all } n \geq N .\]

(2) If the following conditions hold
\[(2.24) \quad X_N \leq M_2 \quad \text{a. s. where } M_2 \text{ is a positive constant}\]
and
\[(2.25) \quad |V_n| \leq v_n \quad \text{a. s. for all } n \geq N ,\]

then for any \(\delta > 0\) there exists a constant \(C(\delta) > 0\) such that
\[(2.26) \quad P[X_n \leq C(\delta) v_n^{\lambda} \text{ for all } n \geq N] > 1 - \delta .\]

**Proof.** (2) is essentially same as the one proved in [14]. Then we shall prove (1).

Taking the expectation on both sides of (2.20), from (2.22) we have
\[(2.27) \quad E[X_{n+1}] \leq (1-a_{n+1})E[X_n] + M_1 \cdot v_{n+1} \quad \text{for all } n \geq N .\]

Putting \(A_n = (E[X_n] v_n^{\lambda})\), from (2.27) we have
\[(2.28) \quad A_n \leq (1-a_{n+1})\left(\frac{v_n}{v_{n+1}}\right)^{\lambda} A_n + M_1 \cdot v_n^{\lambda-1} \quad \text{for all } n \geq N .\]

Hence from (2.17) we have
\[(2.29) \quad A_n \leq A_n + M_1 \cdot v_n^{\lambda-1} \quad \text{for all } n \geq N .\]

Summing up both sides of (2.29) from \(n = N\) to \(N + m - 1\), we obtain
\[(2.30) \quad A_{N+m} \leq A_N + M_1 \cdot \sum_{n=N}^{N+m-1} v_n^{\lambda-1} \leq C \quad \text{for all } m \geq 1 ,\]

where \(C = \left(A_N + M_1 \cdot \sum_{n=N}^{\infty} v_n^{\lambda-1}\right) .\)

Hence, (1) is proved.

**Remark.** Let \(a_n = n^{-1}\) and \(v_n = n^{-(\alpha+1)} (\alpha > 0)\). Then the conditions (2.17) and (2.18) are satisfied by any \(\lambda\) for which \(0 < \lambda < \min \{\alpha(\alpha+1)^{-1}, (1+\alpha)^{-1}\}\) and \(N \geq \frac{\lambda(1+\alpha)}{1-\lambda(\alpha+1)} .\)
§ 3. \textit{R-case.}

In this section we shall treat the case when the observations are taken from $R$. And we assume, at the application of the Robbins-Monro (RM) stochastic approximation method, that an observed random variable at $x \in R$ depends on time $n$, say $Y_n(x)$. In this case, let us consider the problem of finding the root of the equation $E[Y_n(x)] = 0$ for sufficiently large $n$.

Let us give the following RM procedure. Let $X_1$ be an arbitrary constant and let define $X_2, X_3, \ldots$ by the recurrence relation:

$$X_{n+1} = X_n - a_n Y_n$$

where $\{a_n\}$ is a sequence of positive numbers and $\{Y_n\}$ a sequence of random variables.

Throughout this section, we assume the following assumption remains valid.

\textbf{Assumption I.} (i) $\{Y_n(x)\}$ is a sequence of random variables which depend on parameter $x \in R$ and, for each $n$, the expected value of $Y_n(x)$ exists and is a measurable function on $R$, denoted by

$$E[Y_n(x)] = M_n(x).$$

(ii) In (3.1), $Y_n$ is a random variable whose conditional expectation given $X_1, X_2, \ldots, X_n$ coincides with the conditional expectation $Y_n(X_n)$ given $X_n$ and is independent of $X_1, X_2, \ldots, X_{n-1}$. So that

$$E[Y_n | X_1, X_2, \ldots, X_n] = E[Y_n(X_n) | X_n] = M_n(X_n) \quad a.s.$$  

(iii) The sequence $\{a_n\}$ of positive numbers satisfies

$$\lim_{n \to \infty} a_n = 0,$$

$$\sum_{n=1}^{\infty} a_n = +\infty.$$

\textbf{Example.} Let us consider the case when $Y_n(x) = f(x, Z_n)$ where $f$ is a real valued measurable function on $R^2$ and $\{Z_n\}$ a sequence of independent random variables. If $Z_1, Z_2, \ldots$ are identically distributed, then $E[Y_n(x)]$ does not depend on $n$. The RM procedure in the usual sense is this case. If $Z_1, Z_2, \ldots$ are not identically distributed, then $E[Y_n(x)]$ depends on $n$. When $Y_n(x) = f_n(x, Z_n)$ where $f_n$ is a real valued measurable function on $R^2$ for each $n$, $E[Y_n(x)]$ also depends on $n$.

\textbf{Theorem 3.1.} Let $\{\theta_n\}$ be a sequence of the roots of the equations $M_n(x) = 0$, $n = 1, 2, \ldots$. Suppose that there exist three sequences of positive numbers $\{\varepsilon_n\}, \{\rho_n\}$ and $\{A_n\}$ satisfying

(i) $(x - \theta_n)M_n(x) > 0$ if $|x - \theta_n| > \varepsilon_n$ for all $n \geq 1$.

(ii) $\inf_{\varepsilon_n < |x - \theta_n| < \varepsilon_n^{-1}} |M_n(x)| \geq \rho_n$ for all $n \geq 1$.

(iii) $|M_n(x)| \leq A_n(1 + |x - \theta_n|)$ for all $x \in R$ and $n \geq 1$.

(iv) $\sum_{n=1}^{\infty} a_n^2 E[(M_n(X_n) - Y_n)^2] < +\infty$.

(v) $\lim_{n \to \infty} \varepsilon_n = 0$, $(1 \geq \varepsilon_n > 0)$, $\lim_{n \to \infty} a_n A_n = 0$, $\sum_{n=1}^{\infty} a_n \rho_n = +\infty$.

and
(vi) \( \sum_{n=1}^{\infty} |\theta_n - \theta_{n+1}| < +\infty \) or there exists a number \( 0 < \delta < 1 \) and an integer \( N \) such that \( \frac{1}{a_n} |\theta_n - \theta_{n+1}| \leq \delta \inf_{x_n \in [-\theta_n, \theta_n]} |M_n(x)| \) and \( \lim_{n \to \infty} |\theta_n - \theta_{n+1}| = 0. \)

Then it holds that

\[
(3.6) \quad \lim_{n \to \infty} E|X_n - \theta_n|^2 = 0
\]

and

\[
(3.7) \quad \lim_{n \to \infty} |X_n - \theta_n| = 0 \quad \text{a.s.}
\]

**Remarks:**

(i). In this theorem we do not assume the equation \( M_n(x) = 0 \) has a unique root. V. Dupac [3] assumed that the equation has a unique root. T. T. Young and R. A. Westerberg [12] assumed that the root of the equation \( M_n(x) = 0 \) is not only unique, but does not depend on \( n \).

(ii) Let \( \{a_n\} \) be a given sequence of numbers. Consider the problem of finding the root of the equation \( E[Y_n(x)] = a_n \), where \( \{Y_n(x)\} \) is a sequence of random variables. Putting \( Y_n(x) = \tilde{Y}_n(x) - a_n \), then, the problem turns out to be same as the one stated at the beginning of this section.

(iii) In Theorem 3.1, we can omit the conditions (3.4) and (3.5).

**Proof.** Without loss of generality we can assume that

\[
0 < a_n \leq 1 \quad \text{and} \quad 0 < a_n A_n \leq 1 \quad \text{for all} \quad n \leq 1.
\]

Define the measurable transformation in Lemma 1 by

\[
T_n(r) = r + (\theta_n - \theta_{n+1}) - a_n (r + \theta_n) \quad \text{for} \quad n \geq 1 \quad \text{and} \quad r \in R.
\]

And put

\[
U_n = a_n (M_n(X_n) - Y_n) \quad \text{for} \quad n \geq 1.
\]

From Assumption I and (iv), we have

\[
E[U_n | X_n, \ldots, X_1] = 0 \quad \text{a.s. for} \quad n \geq 1,
\]

\[
\sum_{n=1}^{\infty} E|U_n|^2 < +\infty.
\]

From (3.1), (3.9) and (3.10) we have

\[
X_n + 1 - \theta_{n+1} = T_n(X_n - \theta_n) + U_n \quad \text{for} \quad n \geq 1.
\]

Putting \( Z_n = X_n - \theta_n \) we have

\[
Z_{n+1} = T_n(Z_n) + U_n \quad \text{for all} \quad n \geq 1.
\]

We shall show \( T_n \) satisfies the conditions in Lemma 1. Then (3.6) and (3.7) follow by Lemma 1. First, let us assume \( |r| \leq \varepsilon_n \). Then, from (3.8), (iii) and (v) we have

\[
|T_n(r)| \leq |r| + |\theta_n - \theta_{n+1}| + a_n |M_n(r + \theta_n)|
\]

\[
= \varepsilon_n + |\theta_n - \theta_{n+1}| + 2a_n A_n \quad \text{for all} \quad n \geq 1.
\]

Second, let us assume \( \varepsilon_n < |r| < \varepsilon_n^{-1} \). Noting (3.9), \( r M_n(r + \theta_n) > 0 \) and \( \inf_{\varepsilon_n < |r| < \varepsilon_n^{-1}} |M_n(r + \theta_n)| > \rho_n \) we can obtain the inequality
On Robbins-Monro Stochastic Approximation Method

(3.16) \(|T_n(r)| \leq \max \{|\theta_n - \theta_{n+1}| + a_n A_n, |r| - a_n \rho_n + |\theta_n - \theta_{n+1}|\} \) for all \(n \geq 1\).

Third, let us assume \(|r| \geq \varepsilon_n^{-1}\). By the arguments similar to (3.16) we have

(3.17) \(|T_n(r)| \leq \max \{|\theta_n - \theta_{n+1}| + a_n A_n, |r| + |\theta_n - \theta_{n+1}| - a_n M_n(r+\theta_n)|\} \)

for all \(n \geq 1\).

Suppose that \(\sum_{n=1}^{\infty} |\theta_n - \theta_{n+1}| < +\infty\). From (3.15), (3.16) and (3.17), we have

(3.18) \(|T_n(r)| \leq \max \{\varepsilon_n + |\theta_n - \theta_{n+1}| + 2a_n A_n, |r| + |\theta_n - \theta_{n+1}| - \gamma_n(r)\} \)

where

\[ \gamma_n(r) = a_n \rho_n \quad \text{if} \quad |r| < \varepsilon_n^{-1} \]

\[ = 0 \quad \text{if} \quad |r| \geq \varepsilon_n^{-1}. \]

Let \(\{r_n\}\) be a sequence of numbers which satisfies \(\sup |r_n| < \infty\). Then, from (v), there exists a positive integer \(N_0\) which satisfies \(\sup |r_n| < \varepsilon_n^{-1}\) for all \(m \geq N_0\). Hence, we have \(|r_m| < \varepsilon_m^{-1}\) for all \(m \geq N_0\). From (v), (3.18) and the above argument, it follows that \(\sum_{n=1}^{\infty} \gamma_n(r_n) = +\infty\). Thus, if the first condition of (vi) holds, then the conditions (2.1), (2.2) and (2.3) in Lemma 1 are satisfied. Suppose that the second condition of (vi) holds. From (3.15), (3.16), (3.17) and the second condition of (vi), we have

(3.19) \(|T_n(r)| \leq \max \{\varepsilon_n + |\theta_n - \theta_{n+1}| + 2a_n A_n, |r| - (1-\delta) \gamma_n(r)\} \)

for all \(n \geq N\),

where \(\gamma_n(r)\) is defined in (3.18). Hence, if the second condition of (vi) holds, then (2.1), (2.2) and (2.3) are also satisfied. Noting \((a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)\), from (3.1), (iii), (iv) and (v) we can easily show that

(3.20) \(E|X_n - \theta_n|^2 < \infty \quad \text{for all} \quad n \geq 1. \)

Thus all conditions in Lemma 1 are satisfied. This completes the proof.

The following corollary is concerning the case when \(\theta_n\) is the unique root of the equation \(M_n(x) = 0\).

**COROLLARY 3.1.** Let \(\{\theta_n\}\) be the sequence of the roots of the equations \(M_n(x) = 0\), \(n = 1, 2, \ldots\). Suppose that the following conditions are satisfied:

(i) \(M_n(x)(x-\theta_n) \geq 0 \quad \text{for all} \quad x \in \mathbb{R} \quad \text{and} \quad n \geq 1\),

(ii) there exist two positive constants \(C_1\) and \(C_2\) such that

\[ C_1 |x - \theta_n| \leq |M_n(x)| \leq C_2 |x - \theta_n| \quad \text{for all} \quad n \geq 1 \quad \text{and} \quad x \in \mathbb{R}, \]

(iii) the condition (iv) in Theorem 3.1 is satisfied,

(iv) \(\sum_{n=1}^{\infty} |\theta_n - \theta_{n+1}| < +\infty\) or \(\lim_{n \to \infty} \left|\frac{\theta_n - \theta_{n+1}}{a_n}\right| = 0\).

Then (3.6) and (3.7) hold true.

**PROOF.** First, we suppose \(\lim_{n \to \infty} \frac{|\theta_n - \theta_{n+1}|}{a_n} = 0\). Let \(\{\pi_n\}\) be a sequence of positive numbers such that \(\lim_{n \to \infty} \pi_n = 0\) and \(\sum_{n=1}^{\infty} a_n \pi_n = +\infty\), and let \(0 < \delta < 1\).

In Theorem 3.1, we put \(\varepsilon_n = \frac{1}{\delta} \max \left\{\frac{|\theta_n - \theta_{n+1}|}{a_n}, \pi_n\right\}\). Let us assume \(\varepsilon_n < |x - \theta|\).
Then, noting (i) and (ii), we have

\[ M_n(x) \geq C_1 c_n^2 > 0 \quad \text{for all } n \geq 1, \]

and

\[ |M_n(x)| \geq C_2 c_n \quad \text{for all } n \geq 1. \]

Hence, putting \( \rho_n = C_1 c_n \) and \( A_n = C_1 \), we have \( \lim \rho_n = 0, \sum_{n=1}^{\infty} A_n = +\infty \) and \( \inf \delta |M_n(x)| \geq \delta C_1 c_n \geq \frac{|\theta_n - \theta_{n+1}|}{a_n} \) for all \( n \geq N_1 \). Thus all conditions in Theorem 3.1 are satisfied.

Next, we assume \( \sum_{n=1}^{\infty} |\theta_n - \theta_{n+1}| < +\infty \). Putting \( \varepsilon_n = \pi_n \), it is easily seen that the conditions in Theorem 3.1 are also satisfied.

Thus the proof of the corollary is completed.

Next, we shall give some inequalities concerning the rates of convergences.

**Theorem 3.2.** Let \( N \) be a positive integer and \( \{\theta_n\} \) be a sequence of the roots of the equations \( M_n(x) = 0, n = 1, 2, \ldots \). Suppose that the conditions (i) and (ii) in Corollary 3.1.1 are satisfied. Suppose that there exists a sequence of positive numbers \( \{K_n\} \) which satisfies

\[ E \left( (M_n(x) - Y_n)^2 | X_1, \ldots, X_n \right) \leq K_n \quad \text{a.s. for all } n \geq 1. \]

(1) If there exists a number \( 0 < \lambda < 1 \) such that

\[ (1 - C_1 a_n)(v_n/v_{n+1})^2 \leq 1 \quad \text{for all } n \geq N, \]

and

\[ \sum_{n=1}^{\infty} v_n^{n-1} < +\infty, \]

where \( v_n = \max \left\{ \frac{|\theta_n - \theta_{n+1}|^2}{a_n}, a_n^2 K_n \right\} \), then there exists a constant \( C > 0 \) such that

\[ E |X_n - \theta_n|^2 \leq C v_n^2 \quad \text{for all } n \geq 1. \]

(2) If there exists a number \( 0 < \lambda < 1 \) which makes (3.25) hold for all \( n \geq 1 \) and (3.26) hold, and if it is satisfied that

\[ \sup_n a_n < C_1 C_2, \]

then for any \( \delta > 0 \) there exists a constant \( C(\delta) > 0 \) such that

\[ P[|X_n - \theta_n|^2 \leq C(\delta) v_n^2 \quad \text{for all } n \geq 1] > 1 - \delta. \]

**Proof.** From (3.1), Assumption 1, (i), (ii) in Corollary 3.1.1 and (3.24) we have

\[ E[|X_{n+1} - \theta_{n+1}|^2 | X_1, X_2, \ldots, X_n] \]

\[ \leq (1 - 2C_1 a_n + C_2^2 a_n) |X_n - \theta_n|^2 + |\theta_n - \theta_{n+1}|^2 + a_n^2 K_n \]

\[ + 2|\theta_n - \theta_{n+1}| |X_n - \theta_n| + 2C_2 a_n |X_n - \theta_n| |\theta_n - \theta_{n+1}| \quad \text{a.s. for all } n \geq 1. \]

Noting \( 2a b \leq k a^2 + b^2/k \) for any \( k > 0 \), we have
(3.31) $2|\theta_n-\theta_{n+1}| |X_n-\theta_n| \leq \frac{C_1a_n}{k} X_n-\theta_n \geq + k C_1a_n |\theta_n-\theta_{n+1}|^2$

and

(3.32) $2C_2a_n |X_n-\theta_n| |\theta_n-\theta_{n+1}| \leq \frac{C_4a_n^2}{k} X_n-\theta_n \geq + k |\theta_n-\theta_{n+1}|^2$ for all $n \geq 1$.

where $k$ is any positive constant. First we shall prove (1). If we substitute $k=2$ in (3.31) and (3.32), then from (3.30) we get

(3.33) \[ E[|X_{n+1}-\theta_{n+1}|^2 |X_1, \ldots, X_n] \leq \left(1-\frac{3}{2}C_1a_n+\frac{3}{2}C_2a_n^2\right) |X_n-\theta_n|^2 + \left(3+\frac{2}{C_1a_n}\right) |\theta_n-\theta_{n+1}|^2 + a_n^2 K_n \]

a.s. for all $n \geq 1$.

Hence, from (3.4) and (3.33), there exist a positive integer $N_0 \geq N$ and a constant $C_3 \geq 1$ such that

(3.34) \[ E[|X_{n+1}-\theta_{n+1}|^2 |X_1, \ldots, X_n] \leq (1-C_1a_n) |X_n-\theta_n|^2 + C_3v_n \]

a.s. for all $n \geq N_0$.

And noting $E |X_n-\theta_n|^2 < \infty$ for all $n \geq 1$, we have (3.27) by Lemma 3. Next, we shall prove (2). Since $k+1 \downarrow 1$ as $k \to \infty$, from (3.28) we can take $k_0$ which satisfies

(3.35) \[ \frac{C_4}{k_0+1} \geq a_n \quad \text{for all } n \geq 1. \]

In (3.31) and (3.32), we put $k=k_0$. Then we have that there exists a constant $\bar{C}(k_0) \geq 1$ such that

(3.36) \[ E[|X_{n+1}-\theta_{n+1}|^2 |X_1, \ldots, X_n] \leq \left(1-C_1a_n-C_1a_n \left(\frac{1-k_0}{k_0}\right)+C_2a_n^2 \left(\frac{1-k_0}{k_0}\right)\right) |X_n-\theta_n|^2 + \bar{C}(k_0)v_n \]

a.s. for all $n \geq 1$.

From (3.35) we have $-C_1(k_0-1)+C_4(k_0+1)a_n \leq 0$ for all $n \geq 1$. Hence, we have

(3.37) \[ E[|X_{n+1}-\theta_{n+1}|^2 |X_1, \ldots, X_n] \leq \left(1-C_1a_n\right) |X_n-\theta_n|^2 + \bar{C}(k_0)v_n \]

a.s. for all $n \geq 1$.

Noting that $X_1$ is a constant, we can prove (2) by Lemma 3. Thus the proof of the theorem is completed.

**EXAMPLES.** Let $C_1=C_2=1$, $|\theta_n-\theta_{n+1}| = K(n+1)^{-\left(\frac{1+\alpha}{1+\alpha}\right)}$ and $K_n \equiv \text{const}$. And let us put $a_n = (n+1)^{-1}$. From the remark in § 2, if $0 < \alpha < 1$, then putting $0 < \lambda < \frac{\alpha}{1+\alpha}$ we have $E |X_n-\theta_n|^2 \leq C \cdot n^{-2(\alpha+1)}$ for all $n \geq 1$, and putting $0 < \lambda < \min \left\{ \frac{\alpha}{1+\alpha}, \frac{1}{2(\alpha+1)} \right\}$ we have $P[|X_n-\theta_n|^2 \leq C(\delta)n^{-2(\alpha+1)}$ for all $n \geq 1] > 1-\delta$. And if $\alpha \geq 1$, then putting $0 < \lambda < \frac{1}{2}$ we have $E |X_n-\theta_n|^2 \leq C \cdot n^{-2\lambda}$ for all $n \geq 1$, and putting $0 < \lambda < \frac{1}{4}$ we have $P[|X_n-\theta_n|^2 \leq C(\delta)n^{-2\lambda}$ for all $n \geq 1] > 1-\delta$. 


Let $M(x)$ be a real valued measurable (unknown) function on $R$, and let us consider the problem of finding the root of the equation $M(x) = 0$. If a random variable $Y(x)$ satisfying $E[Y(x)] = M(x)$ is available to us at all $x$, then our problem becomes the usual RM method. We shall consider the case when a time-dependent random variable $Y_n(x)$ is given to us at each time $n$. In this case we assume neither that $E[Y(x)]$ coincides with $M(x)$, nor that the equation $E[Y_n(x)] = 0$ has a root. Nevertheless, we shall use the recurrence relation (3.1) also in the subsequent arguments.

**Theorem 3.3.** Let $0$ be a root of the equation $M(x) = 0$. Suppose that there exist three sequences of positive numbers $\{\epsilon_n\}, \{\rho_n\}, \{A_n\}$ satisfying

(i) $(x - \theta)M_n(x) > 0$ if $|x - \theta| > \epsilon_n$ for all $n \geq 1$,

(ii) $\inf_{\epsilon_n < |x - \theta| < \epsilon_n} |M_n(x)| \geq \rho_n$ for all $n \geq 1$,

(iii) $|M_n(x)| \leq A_n(1 + |x - \theta|)$ for all $x \in R$ and $n \geq 1$,

(iv) $\sum_{n=1}^{\infty} a_n^2 E[M_n(X_n) - Y_n] < +\infty$,

(v) $\lim_{n \to \infty} \epsilon_n = 0$ ($0 < \epsilon_n \leq 1$), $\lim_{n \to \infty} \rho_n = 0$, $\sum_{n=1}^{\infty} a_n \rho_n = +\infty$.

Then it holds that

$$\lim_{n \to \infty} E|X_n - \theta|^2 = 0$$

and

$$\lim_{n \to \infty} |X_n - \theta|^2 = 0 \quad \text{a.s.}$$

**Remark.** In this theorem $0$ is the root of the equation $M(x) = 0$, but may not be unique. Moreover, the equation $M_n(x) = 0$ may not have any root.

**Proof.** Let us put

$$T_n(x) = x - \theta - a_n M_n(x)$$

and

$$U_n = a_n (M_n(x) - Y_n).$$

Then we can prove (3.38) and (3.39) by the arguments similar to those in Theorem 3.1.

When the equation $M_n(x) = 0$ has a unique root, we obtain the following corollary.

**Corollary 3.3.1.** Let $\theta$ be a root of the equation $M(x) = 0$ and $\{\theta_n\}$ be the sequence of the roots of $M_n(x) = 0$, $n = 1, 2, \ldots$. Suppose that the conditions (i), (ii) and (iii) in Corollary 3.1.1 are satisfied, and that

$$\lim_{n \to \infty} \epsilon_n = 0.$$

Then (3.6), (3.7), (3.38) and (3.39) hold true.

**Proof.** Putting $A_n = \max \{C_2, C_3 |\theta - \theta_n|\}$ and

$$\epsilon_n = \left(1 + \frac{C_2^2}{C_3^2}\right) |\theta - \theta_n| + \pi_n$$

where $\{\pi_n\}$ is defined in the proof of Corollary 3.1.1 for $n \geq 1$, and noting $|x - \theta| \leq |\theta - \theta_n| + |x - \theta_n|$, then we have $|x - \theta_n| > \frac{C_2^2}{C_3^2} |\theta - \theta_n| + \pi_n$ for
\[|x - \theta| > \varepsilon_n. \] Hence, we have

\[(3.44) \quad (x - \theta)M_n(x) \geq C_1 \pi_n > 0 \quad \text{for} \quad |x - \theta| > \varepsilon_n\]

and

\[(3.45) \quad \inf_{\varepsilon_n < |x - \theta| < \varepsilon_n^{-1}} |M_n(x)| > C_1 \pi_n \quad \text{for all} \quad n \geq 1.\]

The conditions in Theorem 3.3 follow from (3.41) and the inequality \(|M_n(x)| \leq C_2 |x - \theta| + C_2 |\theta - \theta_n|\). (3.6) and (3.7) follow from the inequality \(|X_n - \theta_n| \leq |X_n - \theta| + |\theta - \theta_n|\).

Next, we give the inequalities concerning the rate of convergence that are similar to those in Theorem 3.2.

**Theorem 3.4.** Let \(N\) be a positive integer and \(\theta\) be the root of the equation \(M(x) = 0\). And let \(\{\theta_n\}\) be a sequence of the roots of the equations \(M_n(x) = 0, n = 1, 2, \ldots\). Suppose that the conditions (i) and (ii) in Corollary 3.1.1 are satisfied. Suppose that there exists a sequence of positive numbers \(\{K_n\}\) which satisfy (3.24) for all \(n \geq 1\).

1. If there exists a number \(0 < \lambda < 1\) which satisfies (3.25) for all \(n \geq N\) and (3.26) with \(v_n = \max\{a_n |\theta_n - \theta|^2, a_n^2 K_n\}\), then there exists a constant \(C > 0\) such that

\[(3.46) \quad E[X_n - \theta]^2 \leq C v_n^2 \quad \text{for all} \quad n \geq 1.\]

2. If there exists a number \(0 < \lambda < 1\) which satisfies (3.25) for all \(n \geq 1\) and (3.26) with \(v_n = \max\{a_n |\theta_n - \theta|^2, a_n^2 K_n\}\) and if (3.28) is satisfied, then for any \(\delta > 0\) there exists a constant \(C(\delta) > 0\) such that

\[(3.47) \quad P[|X_n - \theta|^2 \leq C(\delta) v_n^2 \text{ for all } n \geq 1] > 1 - \delta.\]

**Proof.** From (3.1), Assumption I, (i) and (ii) in Corollary 3.1.1 and (3.24) we have

\[(3.48) \quad E[|X_{n+1} - \theta|^2 |X_1, \ldots, X_n]\]

\[\leq |X_n - \theta|^2 + a_n^2 K_n + C_2^2 a_n^2 |X_n - \theta_n|^2 + 2C_2 a_n^2 |X_n - \theta_n| |\theta_n - \theta| - 2C_1 a_n |X_n - \theta_n|^2 \quad \text{a.s. for all} \quad n \geq 1.\]

Noting \(X_n - \theta_n = X_n - \theta + \theta - \theta_n\) and \(C_2 \geq C_1\), we have

\[(3.49) \quad C_2^2 a_n^2 |X_n - \theta_n|^2 \leq \frac{k+1}{k} C_2^2 a_n^2 |X_n - \theta|^2 + (k+1) C_2^2 a_n^2 |\theta_n - \theta|^2\]

and

\[(3.50) \quad 2C_2 a_n |X_n - \theta_n| |\theta_n - \theta| - 2C_1 a_n |X_n - \theta_n|^2 \leq 6C_2 a_n |X_n - \theta| |\theta - \theta_n| + 2C_2 a_n |\theta - \theta_n|^2 - 2C_1 a_n |X_n - \theta|^2 \]

\[\leq \frac{a_n C_2}{k} |X_n - \theta|^2 + \frac{9C_2 a_n}{C_1} |\theta_n - \theta|^2 + 2C_2 a_n |\theta_n - \theta|^2 - 2C_1 a_n |X_n - \theta|^2\]

where \(k\) is a positive number. Therefore, from (3.49) and (3.50), we have

\[(3.51) \quad E[|X_{n+1} - \theta|^2 |X_1, \ldots, X_n]
\]

\[\leq \left[1 - a_n C_1 \left\{C_2 a_n \left(1 - \frac{1}{k}\right) - \frac{k+1}{k} C_2^2 a_n^2 \right\}\right] |X_n - \theta|^2 + \tilde{C}(k)(a_n |\theta_n - \theta| + a_n^2 K_n) \quad \text{a.s. for all} \quad n \geq 1.\]
where $\tilde{C}(k)$ is some positive constant which depends on $k>0$.

First we shall show (1). Substituting $k=2$ in (3.51), by the same arguments as in Theorem 3.2 we obtain (1).

Let $k_0$ be a positive number satisfying $\frac{k_0-1}{k_0+1} C_1/C_2 > a_n$ for all $n \geq 1$. By virtue of (2.28) such a number $k_0$ exists. Hence, from (3.51), we have

$$E[|X_{n+1} - \theta|^4 | X_1, \ldots, X_n]$$

$$\leq (1 - a_n C_3) |X_n - \theta|^2 + \tilde{C}(k_0) (a_n |\theta_n - \theta| + a_n^2 K_n) \text{ a.s. for all } n \geq 1,$$

where $\tilde{C}(k_0)$ is a positive constant. Thus, (2) follows from Lemma 3.


In this section we shall try to generalize the observation space to a Hilbert space. Instead of $R$ in § 3, we shall treat the case when observations will be taken from a separable Hilbert space, denoted by $H$.

Let us consider the problem of finding the root of the equation $E[Y_n(x)] = 0$ for sufficiently large $n$, where $Y_n(x)$ is an observed random element in $H$ at $x \in H$ and at time $n$ and $0$ the null element in $H$.

Let define the RM procedure in $H$ as follows. Let $X_1$ be arbitrary fixed element in $H$ and define $X_2, X_3, \ldots$ in accordance with the recurrence relation:

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

where $\{Y_n\}$ is a sequence of random elements and $\{a_n\}$ a sequence of positive numbers.

Assumption II. (i) $\{Y_n(x)\}$ is a sequence of random elements in $H$ depending on a parameter $x \in H$. For each $n$, the expectation of $Y_n(x)$ exists and is $\mathcal{H}$-measurable transformation of $H$ into itself, denoted by

$$E[Y_n(x)] = M_n(x),$$

where the expectation operator $E$ is defined § 2.

(ii) In (4.1), $Y_n$ is a random element in $H$ whose conditional expectation given $X_1, X_2, \ldots, X_n$ exists and coincides with the conditional expectation of $Y_n(X_n)$ given $X_n$ and is independent of $X_1, X_2, \ldots, X_{n+1}$, where the conditional expectation of $Y_n$ is defined in § 2. That is,

$$E[Y_n | X_1, \ldots, X_n] = E[Y_n(X_n) | X_n] = M_n(X_n) \text{ a.s..}$$

(iii) The positive sequence $\{a_n\}$ satisfies

$$\sum_{n=1}^{\infty} a_n = +\infty,$$

$$\sum_{n=1}^{\infty} a_n^2 < +\infty.$$

Throughout this section we assume that Assumption II is always satisfied.

Corresponding to Theorem 3.1, we give the following theorem.

Theorem 4.1. Let $\{\theta_n\}$ be a sequence of the roots of the equations $M_n(x) = 0$, $n = 1, 2, \ldots$. Suppose that there exist three sequences of positive numbers $\{\epsilon_n\}, \{\rho_n\}$ and
On Robbins-Monro Stochastic Approximation Method

\[ \{A_n\} \text{ satisfying} \]

(i) \( (x-\theta_n, M_n(x)) > 0 \text{ if } \|x-\theta_n\|^2 > \varepsilon_n \text{ for } n \geq 1, \]

(ii) \( \inf_{\varepsilon_n \leq x-\theta_n \leq \varepsilon_n} |(x-\theta_n, M_n(x))| \geq \rho_n \text{ for } n \geq 1, \]

(iii) \( \|M_n(x)\|^2 \leq A_n(1+\|x-\theta_n\|^2) \text{ for all } x \in H \text{ and } n \geq 1, \]

(iv) \( \sum_{n=1}^{\infty} a_n^2 \mathbb{E} \|M_n(X_n)-Y_n\|^2 < \infty, \]

(v) \( \sum_{n=1}^{\infty} a_n \rho_n = \infty, \sum_{n=1}^{\infty} a_n^2 A_n < \infty, \lim_{n \to \infty} \varepsilon_n = 0 \) \( (0 < \varepsilon_n \leq 1), \]

(vi) \( \sum_{n=1}^{\infty} a_n^2 \|\theta_n-\theta_{n+1}\|^2 < \infty. \)

Then it holds that

\[ \lim_{n \to \infty} \mathbb{E} \|X_n-\theta_n\|^2 = 0, \]

and

\[ \lim_{n \to \infty} \|X_n-\theta_n\| = 0 \text{ a. s.} \]

REMARK. (i) \( \theta_n \) may not be the unique root of the equation \( M_n(x) = 0. \)

(ii) Let \( \{\alpha_n\} \) be a given sequence of elements in \( H. \) And we consider the problem of finding the root of the equation \( \mathbb{E}Y'_n(x) = \alpha_n, \) where \( \{Y'_n(x)\} \) is a sequence of observed random elements in \( H. \) Putting \( Y'_n(x) = Y'_n(x)-\alpha_n, \) the problem can be reduced to the case when \( \alpha_n = 0. \)

PROOF. Without loss of generality we can assume that

\[ 0 < a_n A_n \leq 1, \quad 0 < a_n \leq 1 \text{ for all } n \geq 1. \]

Define the measurable transformation \( T_n \) by

\[ T_n(h) = h + \theta_n - \theta_{n+1} - a_n M(h+\theta_n) \text{ for } h \in H. \]

And we put

\[ U_n = a_n(M(X_n)-Y_n). \]

From (4.1), (4.9) and (4.10) we have

\[ X_{n+1} - \theta_{n+1} = T_n(X_n - \theta_n) + U_n \text{ for all } n \geq 1. \]

Putting \( Z_n = X_n - \theta_n, \) we have

\[ Z_{n+1} = T_n(Z_n) + U_n \text{ for all } n \geq 1. \]

And from (iv) and Assumption II we have

\[ \sum_{n=1}^{\infty} \mathbb{E} \|U_n\|^2 < \infty \]

and

\[ \mathbb{E}[U_n | Z_1, Z_2, \ldots, Z_n] = 0 \text{ a. s. for all } n \geq 1. \]

We show that \( T_n \) satisfies all conditions in Lemma 2, then (4.6) and (4.7) follow from Lemma 2. From (4.9) we have
\( (4.15) \quad \|T_n(h)\|^2 \leq \|h\|^2 + \|\theta_n - \theta_{n+1}\|^2 + a_n^2 \|M_n(h + \theta_n) - 2\|\theta_n - \theta_{n+1}\|\|h\| \quad \text{for all } n \geq 1. \)

Noting \[ 2\|h\| \|\theta_n - \theta_{n+1}\| \leq a_n^2 \|h\|^2 + \frac{1}{a_n^2} \|\theta_n - \theta_{n+1}\|^2 \]
and
\[ 2a_n\|\theta_n - \theta_{n+1}\||M_n(h + \theta_n)|| \leq \|\theta_n - \theta_{n+1}\|^2 + a_n^2 \|M_n(h + \theta_n)\|^2, \]
from (iii) we have
\( (4.16) \quad \|T_n(h)\|^2 \leq (1 + a_n^2 + 2a_n^2 A_n)\|h\|^2 + 2a_n^2 A_n + \left(1 + \frac{1}{a_n^2}\right)\|\theta_n - \theta_{n+1}\|^2 - 2a_n\rho_n \quad \text{for all } h \in H \text{ and } n \geq 1. \)

Suppose \( \varepsilon_n < \|h\|^2 < \varepsilon_n^{-1}. \) From (ii) and (4.16) we have
\( (4.17) \quad \|T_n(h)\|^2 \leq (1 + a_n^2 + 2a_n^2 A_n)\|h\|^2 + 2a_n^2 A_n + \left(1 + \frac{1}{a_n^2}\right)\|\theta_n - \theta_{n+1}\|^2 - 2a_n\rho_n \quad \text{for all } h \in H \text{ and } n \geq 1. \)

Suppose \( \varepsilon_n \geq \|h\|^2. \) Noting \( 2a_n\|\theta_n - \theta_{n+1}\|\|h\| \leq a_n\|h\|^2 + A_n(1 + \|h\|^2), \) we have
\( (4.18) \quad \|T_n(h)\|^2 \leq (1 + a_n^2 + a_n^2 + 2a_n^2 A_n)\|h\|^2 + 2a_n^2 A_n + \left(1 + \frac{2a_n^2}{a_n^2}\right)\|\theta_n - \theta_{n+1}\|^2 \quad \text{for all } n \geq 1. \)

Next, suppose that \( \varepsilon_n^{-1} \leq \|h\|^2, \) then we have
\( (4.19) \quad \|T_n(h)\|^2 \leq (1 + a_n^2 + 2a_n^2 A_n)\|h\|^2 + 2a_n^2 A_n + \left(1 + \frac{2a_n^2}{a_n^2}\right)\|\theta_n - \theta_{n+1}\|^2 \quad \text{for all } n \geq 1. \)

Hence, from (4.17), (4.18), (4.19) and (4.8), we have
\( (4.20) \quad \|T_n(h)\|^2 \leq \max \left\{ 6\varepsilon_n + 2a_n^2 A_n + \frac{3}{a_n^2}\|\theta_n - \theta_{n+1}\|^2, (1 + a_n^2 + 2a_n^2 A_n)\|h\|^2 + a_n^2 A_n + \frac{3}{a_n^2}\|\theta_n - \theta_{n+1}\|^2 \right\} \)
where
\( (4.21) \quad \gamma_n(h) = -2a_n\rho_n \quad \text{if } \|h\|^2 < \varepsilon_n^{-1} \)
\[ = 0 \quad \text{if } \|h\|^2 \geq \varepsilon_n^{-1}. \]

It is easily seen that the same arguments as in Theorem 3.1 yield that \( T_n \) satisfies all conditions in Lemma 2. Thus the proof of the theorem is completed.

The following corollary is about the case when \( \theta_n \) is the unique root of the equation \( M_n(x) = 0. \)

**Corollary 4.1.1.** Suppose that the following conditions are satisfied:

(i) \( M_n(x), x - \theta_n \) \( \geq 0 \) for all \( x \in H \) and \( n \geq 1, \)

(ii) there exist two positive constants \( C_1 \) and \( C_2 \) such that
\[ C_1\|x - \theta_n\|^2 \leq (M_n(x), x - \theta_n), \]
\[ \|M_n(x)\| \leq C_2\|x - \theta_n\| \quad \text{for all } x \in H \text{ and } n \geq 1, \]
(iii) \( \sum_{n=1}^{\infty} a_n E[\|M_n(X_n) - Y_n\|^2] < \infty \),
and
(iv) \( \sum_{n=1}^{\infty} \frac{1}{a_n^2} \|\theta_n - \theta_{n+1}\|^2 < \infty \).

Then (4.6) and (4.7) hold true.

Proof. In Theorem 4.1, we put \( \varepsilon_n = \varepsilon_{n-1} \), where \( \{\varepsilon_n\} \) was defined in the proof of Corollary 3.1.1, and \( A_n = C_2 \). If \( \|x - \theta_n\|^2 > \varepsilon_n \), then it holds that
\[
(M_n(x), x - \theta_n) \geq C_1 \|x - \theta_n\|^2 > C_1 \varepsilon_n > 0.
\]
By virtue of (4.4) and (4.5), (4.6) and (4.7) follow from Theorem 4.1.

Next, we shall give the inequalities concerning the rates of convergences in \( H \).

Theorem 4.2. Let \( N \) be a positive integer and \( \{\theta_n\} \) a sequence of the roots of the equations \( M_n(x) = 0 \), \( n = 1, 2, \ldots \). Suppose that the conditions (i) and (ii) in Corollary 4.1.1 are satisfied. Suppose that there exists a sequence of positive numbers \( \{K_n\} \) such that
\[
(4.22) \quad E[\|M_n(X_n) - Y_n\|^2 | X_1, \ldots, X_n] \leq K_n \quad a.s. \quad \text{for all } n \geq 1.
\]

(1) If there exists a number \( 0 < \lambda < 1 \) satisfying (3.25) for all \( n \geq N \) and (3.26) for
\[ v_n = \max \left\{ \frac{\|\theta_n - \theta_{n+1}\|^2}{a_n}, a_n^2 K_n \right\} \]
then there exists a constant \( C > 0 \) such that
\[
(4.23) \quad E[\|X_n - \theta_n\|^2] \leq C v_n^4 \quad \text{for all } n \geq 1.
\]

(2) If there exists a number \( 0 < \lambda < 1 \) satisfying (3.25) for all \( n \geq 1 \) and (3.26)
with \( v_n = \max \left\{ \frac{\|\theta_n - \theta_{n+1}\|^2}{a_n}, a_n^2 K_n \right\} \), and if (3.28) is satisfied, then for any \( \delta > 0 \) there
exists a constant \( C(\delta) > 0 \) such that
\[
(4.24) \quad P[\|X_n - \theta_n\|^2 \leq C(\delta) v_n^4 \text{ for all } n \geq 1] > 1 - \delta.
\]

Proof. From (4.1) and Assumption II we have
\[
(4.25) \quad E[\|X_{n+1} - \theta_{n+1}\|^2 | X_1, \ldots, X_n] = \|X_n - \theta_n\|^2 + \|\theta_n - \theta_{n+1}\|^2 + a_n^2 E[\|M_n(X_n) - Y_n\|^2 | X_1, \ldots, X_n] + a_n^2 \|M_n(X_n)\|^2 - 2a_n(X_n - \theta_n, M_n(X_n)) + 2(X_n - \theta_n, \theta_n - \theta_{n+1}) + 2a_n(M_n(X_n), \theta_n - \theta_{n+1}) \quad a.s. \quad \text{for all } n \geq 1.
\]
Hence, by the same arguments as in Theorem 3.2 we get (1) and (2) in the theorem.

Let \( M(x) \) be a (unknown) measurable transformation of \( H \) into itself. As we did
in § 3, we consider the problem of finding the root of the equation \( M(x) = 0 \) which is
independent of the time \( n \).

Let us assume that in place of a random element \( Y(x) \) in \( H \) satisfying \( E[Y(x)] = M(x) \), as the observation to \( M(x) \) a time-dependent random element \( Y_n(x) \) which
satisfies Assumption II is given to us at each time \( n \). In this case we assume neither that \( E[Y_n(x)] \) (say \( M_n(x) \)) coincides with \( M(x) \), nor that the equation \( M_n(x) = 0 \) has a root. Under these situations, however, we shall follow the procedure (4.1).

Theorem 4.3. Let \( \theta \) be the root of the equation \( M(x) = 0 \). Suppose that there exist three sequences \( \{\varepsilon_n\}, \{p_n\}, \{A_n\} \) satisfying
\[(i) \ (x-\theta, M_n(x)) > 0 \text{ if } \|x-\theta\|^2 > \varepsilon_n \text{ for all } n \geq 1, \]

\[(ii) \ \inf_{\varepsilon < \|x-\theta\|^2 < \varepsilon_n} (x-\theta, M_n(x)) \geq \rho_n \text{ for all } n \geq 1, \]

\[(iii) \ \|M_n(x)\|^2 \leq A_n(1 + \|x-\theta\|^2) \text{ for all } x \in H \text{ and } n \geq 1, \]

\[(iv) \ \sum_{n=1}^{\infty} \frac{a_n}{\varepsilon_n} \mathbb{E} \|M_n(X_n) - Y_n\|^2 < \infty \]

and

\[(v) \ \lim_{n \to \infty} \varepsilon_n = 0 \quad (0 < \varepsilon_n \leq \varepsilon_1), \quad \sum_{n=1}^{\infty} \frac{a_n}{\varepsilon_n} = \infty, \quad \sum_{n=1}^{\infty} a_n^2 A_n < \infty. \]

Then it holds that

\[(4.26) \ \lim_{n \to \infty} \mathbb{E} \|X_n - \theta\|^2 = 0 \]

and

\[(4.27) \ \lim_{n \to \infty} \|X_n - \theta\| = 0 \quad \text{a.s.} \]

**PROOF.** Let us put \(T_n(h) = h - \theta - a_n M_n(h)\) and \(U_n = a_n (M_n(X_0 \cdot \pi_n) - Y_n)\). Then the proof follows the line similar to the one of Theorem 4.1, and so is omitted.

In the case when the equation \(M_n(x) = 0\) has the unique root, we obtain the following corollary.

**COROLLARY 4.3.1.** Let \(\theta\) be the root of the equation \(M(x) = 0\) and \(\{\theta_n\}\) the sequence of the roots of the equations \(M_n(x) = 0, \ n = 1, 2, \ldots\). Suppose that the conditions (i), (ii) and (iii) in Corollary 4.1.1 are satisfied and that

\[(4.28) \ \lim_{n \to \infty} \|\theta_n - \theta\|^2 = 0. \]

Then (4.6), (4.7), (4.26) and (4.27) hold.

**PROOF.** Putting \(\varepsilon_n = \left(2 + \frac{2C_\varepsilon}{\nu_n}\right)\|\theta_n - \theta\|^2 + \frac{\varepsilon_n}{2}\), where \(\{\pi_n\}\) was defined in the proof of Corollary 3.1.1, and \(A_n = \max\{2C_\varepsilon\|\theta_n - \theta\|^2, 2C_\varepsilon\}\), we get (4.30) and (4.31) by Theorem 4.1. And noting \(\|X_n - \theta\|^2 \leq 2\|X_n - \theta\|^2 + 2\|\theta_n - \theta\|^2\), we get (4.6) and (4.7).

We also obtain the inequalities concerning the rates of convergences.

**THEOREM 4.4.** Let \(N\) be a positive integer and \(\{\theta_n\}\) a sequence of the roots of the equations \(M_n(x) = 0, \ n = 1, 2, \ldots\). Let \(\theta\) be the root of the equation \(M_n(x) = 0\). Suppose that the conditions (i) and (ii) in Corollary 4.1.1 are satisfied. Suppose that there exists a sequence of positive numbers \(\{K_n\}\) which satisfy (4.22) for all \(n \geq 1\).

1. If there exists a number \(0 < \lambda < 1\) which satisfies (3.25) for all \(n \geq N\) and (3.26) with \(\nu_n = \max\{a_n\|\theta_n - \theta\|^2, a_n^2 K_n\}\), then there exists a constant \(C > 0\) such that

\[(4.29) \ \mathbb{E} \|X_n - \theta\|^2 \leq C \nu_n^4 \quad \text{for all } n \geq 1. \]

2. If there exists a number \(0 < \lambda < 1\) which satisfies (3.25) for all \(n \geq 1\) and (3.26) with \(\nu_n = \max\{a_n\|\theta_n - \theta\|^2, a_n^2 K_n\}\) and if (3.28) is satisfied, then for any \(\delta > 0\) there exists a constant \(C(\delta) > 0\) such that

\[(4.30) \ \mathbb{P}[\|X_n - \theta\|^2 \leq C(\delta) \nu_n^4] > 1 - \delta. \]

**PROOF.** From (4.1) and Assumption II we have
On Robbins-monro Stochastic Approximation Method

\[(4.31)\quad E[\|X_{n+1}-\theta\|^2|X_1, \cdots, X_n]\]

\[=\|X_n-\theta\|^2-2a_n(X_n-\theta, M_n(X_n))
+a_n^2\|M_n(X_n)\|^2+a_n^2E[\|M_n(X_n)-Y_n\|^2|X_1, \cdots, X_n]\quad a.s.\quad \text{for all } n \geq 1.\]

Hence, by the same arguments as in Theorem 3.4 we get (1) and (2).

§ 5. Applications.

In this section we shall give some applications of the results which we have obtained in § 4.

(1) Let $M(x)$ be a real valued measurable function on $H$, which is unknown to us. Then we consider the problem of finding the root of the equation $M(x)=0$. Let $\{\tilde{Y}_n(x)\}$ be a sequence of random variable with parameter $x \in H$. We put

\[(5.1)\quad E\tilde{Y}_n(x)=M_n(x)\quad \text{for } n \geq 1 \text{ and } x \in H.\]

In this problem, we follow the procedure:

\[(5.2)\quad X_1 \text{ is an arbitrary fixed element in } H, \quad X_{n+1}=X_n-a_n\tilde{Y}_ne \quad \text{for } n \geq 1,\]

where $e$ is an element in $H$ which satisfies $\|e\|=1$ and $\{a_n\}$ is a sequence of positive numbers which satisfy (4.4) and (4.5). Putting $Y_n(x)=\tilde{Y}_n(x)e$ and $Y_n=\tilde{Y}_ne$, this problem is same as the problem which we gave in § 4.

(2) Density estimation problem ([7], [11]). Consider a sequence $Z_1, Z_2, \cdots$ of independent identically distributed random variables having a unknown probability density function $f$ w.r.t. the Lebesgue measure on $R$. Let us assume $f \in L^2(R)$ where $L^2(R)$ denote the space of all Lebesgue square-integrable functions defined on $R$. Let $K$ be a non-negative valued measurable function on $R$ such that

\[(5.3)\quad \sup_y K(y) < \infty,\]

\[(5.4)\quad \int_R K(y)dy=1,\]

\[(5.5)\quad \lim_{y \to \infty} |y|K(y)=0.\]

From (5.3) and (5.4), we have $K \in L^2(R)$. Let us put

\[(5.6)\quad K_n(x, z) = \frac{1}{h_n}K\left(\frac{x-z}{h_n}\right),\]

where $\{h_n\}$ is a sequence of positive numbers such that $h_n \downarrow 0$, and put

\[(5.7)\quad \hat{f}_n(x) = E[K_n(x, Z_{n+1})].\]

Then we have $K_n(x, z) \in L^2(R)$ for each $z \in R$ and $\hat{f}_n \in L^2(R)$.

In order to estimate $f$, we shall give the following recursive procedure:

\[(5.8)\quad \hat{f}_0 = 0, \quad \hat{f}_{n+1}(y) = \hat{f}_n(y) - a_n\{f_n(y) - K_n(y, Z_{n+1})\} \quad \text{for } n \geq 1 \text{ and } y \in R,\]
where \( \{a_n\} \) is a sequence satisfying (4.4) and (4.5). In § 4, we put \( H = L^2(R) \), \( X_n = f_n \), \( Y_n = f_n - K_n(\cdot, Z_{n+1}) \), \( Y_n(f) = f - K_n(\cdot, Z_{n+1}) \), \( \theta = \hat{\theta} \) and \( \theta_n = \hat{\theta}_n \), then we have

\[
E[Y_n(f)] = f - \hat{f}_n \quad \text{for any } f \in L^2(R),
\]

\[
E[Y_n|X_1, \ldots, X_n] = E[f_n - K_n(\cdot, Z_{n+1})|f_1, \ldots, f_n]
= f_n - \hat{f}_n \quad \text{a.s.}
\]

By (5.3), (5.4) and (5.5) it holds that \( \lim_{n \to \infty} \| \hat{f}_n - \hat{f} \| = 0 \). Hence, from Theorem 4.2, we have \( \lim_{n \to \infty} \| f_n - \hat{f} \| = 0 \) a.s. and \( \lim_{n \to \infty} E\| f_n - \hat{f} \|^2 = 0 \).

Thus, the above problem is regarded as the RM stochastic approximation in the sense of § 4.

REMARK. The above problem is also considered as the RM in \( R \). In this case, if \( \hat{f} \) is a continuous function on \( R \), then we can show that \( \lim_{n \to \infty} E|f_n(y) - \hat{f}(y)| = 0 \) for all \( y \) and that \( \lim_{n \to \infty} |f_n(y) - \hat{f}(y)| = 0 \) a.s. for all \( y \) (\cite{7}).

(3) The learning problem for pattern classification (\cite{9}, \cite{13}, \cite{14}, \cite{15}). Let us consider the problem of estimating an unknown discriminant function \( \hat{D}(y) = q_A f_n(y) - q_B f_n(y) \), where \( (q_A, q_B) \) is a priori distribution on \( \{A, B\} \), and \( f_A \) and \( f_B \) are the probability density functions on \( R \) under the categories \( A \) and \( B \) respectively (\cite{13}, \cite{14}). Let \( \{(Z_n, \Theta_n)\} \) be a training sequence which is independently and identically distributed. We construct the learning algorithm in the manner similar to (5.8):

\[
D_0 = 0,
\]

\[
D_{n+1}(y) = D_n(y) - a_{n+1} \{ D_{n+1}(y) - \rho(\Theta_{n+1}) K_n(y, Z_{n+1}) \} \quad \text{for each } y \in R,
\]

where

\[
\rho(\Theta_n) = 1 \quad \text{if } \Theta_n = A,
= -1 \quad \text{if } \Theta_n = B.
\]

Let us put \( H = L^2(R) \), \( X_n = D_n \), \( Y_n = D_n - \rho(\Theta_{n+1}) K_n(\cdot, Z_{n+1}) \), \( Y_n(D) = D - \rho(\Theta_{n+1}) K_n(\cdot, Z_{n+1}) \) for all \( D \in L^2(R) \), \( \theta = \hat{D} \) and \( \theta_n = \hat{D}_n \) where \( D_n(y) = E[\rho(\Theta_{n+1}) K_n(y, Z_{n+1})] \) for \( y \in R \). Then we also have

\[
E[Y_n(D)] = D - \hat{D} \quad \text{for each } D \in L^2(R) \text{ and } n \geq 1,
\]

\[
E[Y_n|X_1, \ldots, X_n] = E[D_n - \rho(\Theta_{n+1}) K_n(\cdot, Z_{n+1})|D_1, \ldots, D_n]
= D_n - \hat{D}_n \quad \text{a.s.} \quad \text{for } n \geq 1.
\]

Thus, this problem also can be considered as the RM stochastic approximation method in \( H \). And if \( \| \hat{D}_n - \hat{D} \|^2 \to 0 \) as \( n \to \infty \), then the convergences follows from the theorem in § 4.

§ 6. Acknowledgement.

The author is deeply indebted to Professor N. Furukawa of Kyushu University for helpful advices and encouragements.
References


