ON ROBBINS-MONRO STOCHASTIC APPROXIMATION METHOD WITH TIME VARING OBSERVATIONS

Watanabe, Masafumi
Department of Applied Mathematics, Fukuoka University

https://doi.org/10.5109/13096

出版情報：統計数理研究. 16 (3/4), pp.73-91, 1975-03. Research Association of Statistical Sciences
バージョン：
権利関係：

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 descriebed 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 $\mathrm{E}[Y(x)]$ (regression function), is unknown to us. Assuming that the equation $\mathrm{E}[Y(x)]=0$ has the unique root $x=\theta$, it is desired to estimate θ on the basis of observed values $\{Y(x_n)\}$ at the points $x_1, x_2, \cdots$ which are produced by the following recurrence relation

$$\begin{aligned} &x_1 \equiv \text{const.} \\ &x_{n+1}=x_n-a_n y_n\,, \end{aligned} \tag{1.1}$$

where y_n is a random variable whose conditional distribution given $(x_1, x_2, \cdots, x_n)$ coincides with the distribution of random variable $Y(x_n)$ and $\{a_n\}$ 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 $\mathrm{E}[Y_n(x)]=0$ for sufficient large n. Regarding this problem, V. Dupac [3] and T. V. Young and R. A. Westerberg [12] assumed that the equation $\mathrm{E}[Y_n(x)]=0$ 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 $\mathrm{E}[Y(x)]=M(x)$, we can observe, at each time n, a random variable $Y_n(x)$ for which the equation $\mathrm{E}[Y_n(x)]=M(x)$ 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 $\mathrm{E}[Y_n(x)]=0$ 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,

* Department of Applied Mathematics, Fukuoka University, Fukuoka.

the density estimation problems ([7], [11]) and the learning problems for a pattern clasification ([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 Euclidian 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 Euclidian space, R^n the n-dimensional Euclidian 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 $\mathcal{H}$ be the σ-field of subsets of H generated by the open sets. Let $(\Omega, \mathcal{A}, P)$ be a probability space. A random element X in H is a measurable mapping from $(\Omega, \mathcal{A})$ into $(H, \mathcal{H})$.

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 $\mathrm{E}\|X\|<\infty$, then $\mathrm{E}X$ is defined by the requirement $\mathrm{E}(h, X)=(h, \mathrm{E}X)$ for all h in H. Similarly, let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$, then the conditional expectation of X given $\mathcal{B}$, denoted by $\mathrm{E}[X|\mathcal{B}]$, is defined by the requirement $\mathrm{E}[(h, X)|\mathcal{B}]=(h, \mathrm{E}[X|\mathcal{B}])$ a. s. for all h in H. The expectation operator and the coditional 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, $n=1,2,\cdots$, be measurable Transformations of R^n into R. Let $\{\alpha_n\}$, $\{\beta_n\}$, $\{\delta_n\}$ be sequences of non-negative numbers and γ_n be a non-negative valued measurable function on R^n. Suppose that the following conditions are satisfied:*

$$T_n(r_1, \cdots, r_n) \leqq \max\{\alpha_n, (1+\beta)|r_n| - \gamma_n(r_1, \cdots, r_n)+\delta_n\} \tag{2.1}$$

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

$$\lim_{n\to\infty} \alpha_n = 0\,, \qquad \sum_{n=1}^{\infty} \beta_n < +\infty\,, \qquad \sum_{n=1}^{\infty} \delta_n < +\infty \tag{2.2}$$

and

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

uniformly for all sequences $\{r_n\} \subset R$ for which $\sup_n r_n < \infty$. And let X_1 be any constant and $\{U_n\}$ be a sequence of random variables. Define, for each $n \geqq 1$,

$$X_{n+1} = T_n(x_1, \cdots, X_n) + U_n\,. \tag{2.4}$$

Then the conditions, $\mathrm{E}|X_N|^2 < +\infty$, (2.5) $$\sum_{n=1}^{\infty} \mathrm{E}|U_n|^2 < +\infty$$

and

(2.6) $$\mathrm{E}[U_n | X_1, \cdots, X_n] = 0 \quad a.s. \quad \text{for all } n \geqq N$$

imply

(2.7) $$\lim_{n\to\infty} \mathrm{E}|X_n|^2 = 0$$

and

(2.8) $$\lim_{n\to\infty} X_n = 0 \quad a.s.$$

The following lamma 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 γ_n be a non-negative valued measurable function on H^n. Suppose that the following conditions are satisfied:*

(2.9) $$\|T_n(h_1, \cdots, h_n)\|^2 \leqq \max\{\alpha_n, (1+\beta_n)\|h_n\|^2 - \gamma_n(h_1, \cdots, h_n) + \delta_n\}$$

for all $(h_1, \cdots, h_n) \in H^n$ and all $n \geqq N$,

(2.10) $$\lim_{n\to\infty} \alpha_n = 0, \quad \sum_{n=1}^{\infty} \beta_n < +\infty, \quad \sum_{n=1}^{\infty} \delta_n < +\infty$$

and

(2.11) $$\sum_{n=1}^{\infty} \gamma_n(h_1, \cdots, h_n) = +\infty$$

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 $n \geqq 1$,

(2.12) $$X_{n+1} = T_n(X_1, \cdots, X_n) + U_n .$$

Then the conditions, $\mathrm{E}\|X_N\|^2 < \infty$,

(2.13) $$\sum_{n=1}^{\infty} \mathrm{E}\|U_n\|^2 < +\infty$$

and

(2.14) $$\mathrm{E}[U_n | X_1, \cdots, X_n] = 0 \quad a.s. \quad \text{for all } n \geqq N$$

imply

(2.15) $$\lim_{n\to\infty} \mathrm{E}\|X_n\|^2 = 0$$

and

(2.16) $$\lim_{n\to\infty} \|X_n\| = 0 \quad a.s.$$

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 σ-fields, $\mathcal{A}_n \subset \mathcal{A}_{n+1} \subset \mathcal{A}$, where X_n is measurable with respect $\mathcal{A}_n$ for each $n \geqq 1$. And let N be a positive integer and $\{a_n\}$, $\{v_n\}$ be two sequences of positive numbers which satisfy there exists a number $0 < \lambda < 1$ such that*

(2.17) $$(1-a_{n+1})(v_n/v_{n+1})^{\lambda} \leqq 1 \quad \textit{for all } n \geqq N$$

and

(2.18) $$\sum_{n=1}^{\infty} v_n^{1-\lambda} < +\infty .$$

Suppose that the following conditions are satisfied

(2.19) $$0 \leqq X_n \quad \textit{a. s.} \quad \textit{for all } n \geqq 1$$

(2.20) $$\mathrm{E}[X_{n+1} \mid \mathcal{A}_n] \leqq (1-a_{n+1})X_n + M_1 \cdot V_{n+1} \quad \textit{a. s.} \quad \textit{for all } n \geqq 1$$

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

(1) *If the following conditions hold*

(2.21) $$\mathrm{E}X_N < \infty$$

and

(2.22) $$\mathrm{E}|V_n| \leqq v_n \quad \textit{for all } n \geqq N$$

then there exists a constant $C>0$ such that

(2.23) $$\mathrm{E}X_n \leqq C \cdot v_n^{\lambda} \quad \textit{for all } n \geqq N .$$

(2) *If the following conditions hold*

(2.24) $$X_N \leqq M_2 \quad \textit{a. s.} \quad \textit{where } M_2 \textit{ is a positive constant}$$

and

(2.25) $$|V_n| \leqq v_n \quad \textit{a. s.} \quad \textit{for all } n \geqq N ,$$

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

(2.26) $$\mathrm{P}[X_n \leqq C(\delta) v_n^{\lambda} \textit{ for all } n \geqq N] > 1-\delta .$$

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

Taking the expectation on the both sides of (2.20), from (2.22) we have

(2.27) $$\mathrm{E}X_{n+1} \leqq (1-a_{n+1})\mathrm{E}X_n + M_1 \cdot v_{n+1} \quad \text{for all } n \geqq N .$$

Putting $A_n = (\mathrm{E}X_n) v_n^{-\lambda}$, from (2.27) we have

(2.28) $$A_{n+1} \leqq (1-a_{n+1}) \left(\frac{v_n}{v_{n+1}} \right)^{\lambda} A_n + M_1 \cdot v_{n+1}^{1-\lambda} \quad \text{for all } n \geqq N .$$

Hence from (2.17) we have

(2.29) $$A_{n+1} \leqq A_n + M_1 \cdot v_{n+1}^{1-\lambda} \quad \text{for all } n \geqq N .$$

Summing up the both sides of (2.29) from $n=N$ to $N+m-1$, we obtain

(2.30) $$A_{N+m} \leqq A_N + M_1 \cdot \sum_{n=N+1}^{N+m} v_n^{1-\lambda} \leqq C \quad \text{for all } m \geqq 1 .$$

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

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 λ for which $0 < \lambda < \min\{\alpha(\alpha+1)^{-1}, (1+\alpha)^{-1}\}$ and $N \geqq \dfrac{\lambda(1+\alpha)}{1-\lambda(\alpha+1)}$.

§ 3. *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 $X_2, X_3, \cdots$ by the recurrence relation:

$$X_{n+1}=X_n-a_nY_n \tag{3.1}$$

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.

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

$$E[Y_n(x)]=M_n(x)\,. \tag{3.2}$$

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

$$E[Y_n \mid X_1, X_2, \cdots, X_n]=E[Y_n(X_n) \mid X_n]=M_n(X_n) \qquad a.\,s. \tag{3.3}$$

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

$$\lim_{n\to\infty} a_n=0\,, \tag{3.4}$$

$$\sum_{n=1}^{\infty} a_n=+\infty\,. \tag{3.5}$$

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 $Z_1, Z_2, \cdots$ are identically distributed, then $E[Y_n(x)]$ does not depend on n. The RM procedure in the usual sense is this case. If $Z_1, Z_2, \cdots$ 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.

THEOREM 3.1. *Let $\{\theta_n\}$ be a sequence of the roots of the equations $M_n(x)=0$, $n=1, 2, \cdots$. 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* $n \geqq 1$.

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

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

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

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

and

(vi) $\sum_{n=1}^{\infty}|\theta_n-\theta_{n+1}|<+\infty$ *or there exist a number* $0<\delta<1$ *and an integer* N *such that* $\frac{|\theta_n-\theta_{n+1}|}{a_n}\leqq\delta \inf_{\varepsilon_n<|x-\theta_n|}|M_n(x)|$ *and* $\lim_{n\to\infty}|\theta_n-\theta_{n+1}|=0$.

Then it hold that

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

and

(3.7) $$\lim_{n\to\infty}|X_n-\theta_n|=0 \qquad a.\ s.$$

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 $\{\alpha_n\}$ be a given sequence of numbers. Consider the problem of finding the root of the eqnation $E[\hat{Y}_n(x)]=\alpha_n$, where $\{\hat{Y}_n(x)\}$ is a sequence of random variables. Putting $Y_n(x)=\hat{Y}_n(x)-\alpha_n$, 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

(3.8) $$0<a_n\leqq 1 \quad \text{and} \quad 0<a_nA_n\leqq 1 \qquad \text{for all } n\leqq 1.$$

Define the measurable transformation in Lemma 1 by

(3.9) $$T_n(r)=r+(\theta_n-\theta_{n+1})-a_nM_n(r+\theta_n) \qquad \text{for } n\geqq 1 \text{ and } r\in R.$$

And put

(3.10) $$U_n=a_n(M_n(X_n)-Y_n) \qquad \text{for } n\geqq 1.$$

From Assumption I and (iv), we have

(3.11) $$E[U_n|X_1,\cdots,X_n]=0 \qquad a.\ s. \qquad \text{for } n\geqq 1,$$

(3.12) $$\sum_{n=1}^{\infty}E|U_n|^2<+\infty.$$

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

(3.13) $$X_{n+1}-\theta_{n+1}=T_n(X_n-\theta_n)+U_n \qquad \text{for } n\geqq 1.$$

Putting $Z_n=X_n-\theta_n$ we have

(3.14) $$Z_{n+1}=T_n(Z_n)+U_n \qquad \text{for all } n\geqq 1.$$

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 $|r|\leqq\varepsilon_n$. Then, from (3.8), (iii) and (v) we have

(3.15) $$\begin{aligned}|T_n(r)|&\leqq|r|+|\theta_n-\theta_{n+1}|+a_n|M_n(r+\theta_n)|\\ &=\varepsilon_n+|\theta_n-\theta_{n+1}|+2a_nA_n \qquad \text{for all } n\geqq 1.\end{aligned}$$

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

(3.16) $$|T_n(r)| \leqq \max\{|\theta_n-\theta_{n+1}|+a_nA_n,\ |r|-a_n\rho_n+|\theta_n-\theta_{n+1}|\} \qquad \text{for all } n \geqq 1.$$

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

(3.17) $$|T_n(r)| \leqq \max\{|\theta_n-\theta_{n+1}|+a_nA_n,\ |r|+|\theta_n-\theta_{n+1}|-a_n|M_n(r+\theta_n)|\} \qquad \text{for all } n \geqq 1.$$

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

(3.18) $$|T_n(r)| \leqq \max\{\varepsilon_n+|\theta_n-\theta_{n+1}|+2a_nA_n,\ |r|+|\theta_n-\theta_{n+1}|-\gamma_n(r)\}$$

where

$$\gamma_n(r)=a_n\rho_n \qquad \text{if} \quad |r|<\varepsilon_n^{-1}$$
$$=0 \qquad \text{if} \quad |r| \geqq \varepsilon_n^{-1}.$$

Let $\{r_n\}$ be a sequence of numbers which satisfies $\sup_n |r_n|<\infty$. Then, from (v), there exists a positive integer N_0 which satisfies $\sup_n |r_n|<\varepsilon_m^{-1}$ for all $m \geqq N_0$. Hence, we have $|r_m|<\varepsilon_m^{-1}$ for all $m \geqq N_0$. 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) $$|T_n(r)| \leqq \max\{\varepsilon_n+|\theta_n-\theta_{n+1}|+2a_nA_n,\ |r|-(1-\delta)\gamma_n(r)\} \qquad \text{for all } n \geqq N.$$

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 $(a+b+c+d)^2 \leqq 4(a^2+b^2+c^2+d^2)$, from (3.1), (iii), (iv) and (v) we can easily show that

(3.20) $$\mathrm{E}|X_n-\theta_n|^2<\infty \qquad \text{for all } n \geqq 1.$$

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

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

COROLLARY 3.1.1. *Let $\{\theta_n\}$ be the sequence of the roots of the equations $M_n(x)=0$, $n=1, 2, \cdots$. Suppose that the following conditions are satisfied:*

(i) *$M_n(x)(x-\theta_n) \geqq 0$ for all $x \in R$ and $n \geqq 1$,*

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

$$C_1|x-\theta_n| \leqq |M_n(x)| \leqq C_2|x-\theta_n| \qquad \textit{for all } n \geqq 1 \textit{ and } x \in R,$$

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

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

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 $\varepsilon_n=\frac{1}{\delta\cdot C_1}\max\left\{\frac{|\theta_n-\theta_{n+1}|}{a_n},\ \pi_n\right\}$. Let us assume $\varepsilon_n<|x-\theta|$.

Then, noting (i) and (ii), we have

(3.21) $$M_n(x)(x-\theta_n)\geqq C_1\varepsilon_n^2>0 \quad \text{for all } n\geqq 1,$$

and

(3.22) $$|M_n(x)|\geqq C_1\varepsilon_n \quad \text{for all } n\geqq 1.$$

Hence, putting $\rho_n=C_1\varepsilon_n$ and $A_n=C_1$, we have $\lim_{n\to\infty}\varepsilon_n=0$, $\sum_{n=1}^{\infty}a_n\rho_n=+\infty$ and $\inf_{|x-\theta_n|<\varepsilon_n}\delta|M_n(x)|\geqq\delta C_1\varepsilon_n\geqq\frac{|\theta_n-\theta_{n+1}|}{a_n}$ for all $n\geqq N_1$. 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 $M_n(x)=0$, $n=1,2,\cdots$. 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*

(3.24) $$\mathrm{E}[(M_n(X_n)-Y_n)^2|X_1,\cdots,X_n]\leqq K_n \quad a.s. \quad \textit{for all } n\geqq 1.$$

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

(3.25) $$(1-C_1a_n)(v_n/v_{n+1})^{\lambda}\leqq 1 \quad \textit{for all } n\geqq N,$$

and

(3.26) $$\sum_{n=1}^{\infty}v_n^{1-\lambda}<+\infty,$$

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

(3.27) $$\mathrm{E}|X_n-\theta_n|^2\leqq Cv_n^{\lambda} \quad \textit{for all } n\geqq 1.$$

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

(3.28) $$\sup_n a_n<C_1/C_2^2,$$

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

(3.29) $$\mathrm{P}[|X_n-\theta_n|^2\leqq C(\delta)v_n^{\lambda} \textit{ for all } n\geqq 1]>1-\delta.$$

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

(3.30) $$\begin{aligned}&\mathrm{E}[|X_{n+1}-\theta_{n+1}|^2|X_1,X_2,\cdots,X_n]\\ &\leqq(1-2C_1a_n+C_2^2a_n)|X_n-\theta_n|^2+|\theta_n-\theta_{n+1}|^2+a_n^2K_n\\ &\quad+2|\theta_n-\theta_{n+1}||X_n-\theta_n|+2C_2a_n|X_n-\theta_n||\theta_n-\theta_{n+1}| \quad a.s. \quad \text{for all } n\geqq 1.\end{aligned}$$

Noting $2ab\leqq ka^2+b^2/k$ for any $k>0$, we have

$$(3.31)\qquad 2|\theta_n-\theta_{n+1}||X_n-\theta_n| \leqq \frac{C_1a_n}{k}|X_n-\theta_n|^2+\frac{k}{C_1a_n}|\theta_n-\theta_{n+1}|^2$$

and

$$(3.32)\qquad 2C_2a_n|X_n-\theta_n||\theta_n-\theta_{n+1}| \leqq \frac{C_2^2a_n^2}{k}|X_n-\theta_n|^2+k|\theta_n-\theta_{n+1}|^2 \qquad \text{for all } n\geqq 1,$$

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

$$(3.33)\qquad \mathrm{E}[|X_{n+1}-\theta_{n+1}|^2|X_1,\cdots,X_n]$$
$$\leqq\left(1-\frac{3}{2}C_1a_n+\frac{3}{2}C_2^2a_n^2\right)|X_n-\theta_n|^2+\left(3+\frac{2}{C_1a_n}\right)|\theta_n-\theta_{n+1}|^2+a_n^2K_n$$
$$a.\,s.\qquad \text{for all } n\geqq 1.$$

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

$$(3.34)\qquad \mathrm{E}[|X_{n+1}-\theta_{n+1}|^2|X_1,\cdots,X_n]\leqq(1-C_1a_n)|X_n-\theta_n|^2+C_3v_n$$
$$a.\,s.\qquad \text{for all } n\geqq N_0.$$

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

$$(3.35)\qquad \frac{C_1}{C_2^2}\cdot\frac{k_0-1}{k_0+1}\geqq a_n \qquad \text{for all } n\geqq 1.$$

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

$$(3.36)\qquad \mathrm{E}[|X_{n+1}-\theta_{n+1}|^2|X_1,\cdots,X_n]$$
$$\leqq\left\{1-C_1a_n-C_1a_n\left(\frac{k_0-1}{k_0}\right)+C_2^2a_n^2\left(\frac{1+k_0}{k_0}\right)\right\}|X_n-\theta_n|^2+\hat{C}(k_0)v_n \qquad a.\,s.$$
$$\text{for all } n\geqq 1.$$

From (3.35) we have $-C_1(k_0-1)+C_2^2(k_0+1)a_n\leqq 0$ for all $n\geqq 1$. Hence, we have

$$(3.37)\qquad \mathrm{E}[|X_{n+1}-\theta_{n+1}|^2|X_1,\cdots,X_n]\leqq(1-C_1a_n)|X_n-\theta_n|^2+\hat{C}(k_0)v_n$$
$$a.\,s.\qquad \text{for all } n\geqq 1.$$

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

EXAMPLES. Let $C_1=C_2=1$, $|\theta_n-\theta_{n+1}|=K(n+1)^{-\left(1+\frac{\alpha}{2}\right)}$ and $K_n\equiv$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 $\mathrm{E}|X_n-\theta_n|^2\leqq C\cdot n^{-\lambda(\alpha+1)}$ for all $n\geqq 1$, and putting $0<\lambda<\min\left\{\frac{\alpha}{1+\alpha},\frac{1}{2(\alpha+1)}\right\}$ we have $\mathrm{P}[|X_n-\theta_n|^2\leqq C(\delta)n^{-\lambda(\alpha+1)}$ for all $n\geqq 1]>1-\delta$. And if $\alpha\geqq 1$, then putting $0<\lambda<\frac{1}{2}$ we have $\mathrm{E}|X_n-\theta_n|^2\leqq C\cdot n^{-2\lambda}$ for all $n\geqq 1$, and putting $0<\lambda<\frac{1}{4}$ we have $\mathrm{P}[|X_n-\theta_n|^2\leqq C(\delta)n^{-2\lambda}$ for all $n\geqq 1]>1-\delta$.

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 $\mathrm{E}[Y(x)]=M(x)$ is available to us at all x, then our problem becomes the usual RM method. We shall consider the case when $Y(x)$ in the above sense is not given to us, but as the observation to $M(x)$ a time-dependent random variable $Y_n(x)$ is given to us at each time n. In this case we assume neither that $\mathrm{E}[Y_n(x)]$ coincides with $M(x)$, nor that the equation $\mathrm{E}[Y_n(x)]=0$ has a root. Nevertheless, we shall use the recurrence relation (3.1) also in the subsequent arguments.

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

(i) $(x-\theta)M_n(x)>0$ *if* $|x-\theta|>\varepsilon_n$ *for all* $n\geqq 1$,

(ii) $\inf\limits_{\varepsilon_n<|x-\theta|<\varepsilon_n^{-1}} |M_n(x)|\geqq \rho_n$ *for all* $n\geqq 1$,

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

(vi) $\sum\limits_{n=1}^{\infty} a_n^2 \mathrm{E}|M_n(X_n)-Y_n|^2<+\infty$,

(v) $\lim\limits_{n\to\infty}\varepsilon_n=0$ $(0<\varepsilon_n\leqq 1)$, $\lim\limits_{n\to\infty} a_nA_n=0$, $\sum\limits_{n=1}^{\infty} a_n\rho_n=+\infty$.

Then it holds that

$$\lim_{n\to\infty}\mathrm{E}|X_n-\theta|^2=0 \tag{3.38}$$

and

$$\lim_{n\to\infty}|X_n-\theta|^2=0 \qquad a.\,s. \tag{3.39}$$

REMARK. In this theorem θ 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

$$T_n(x)=x-\theta-a_nM_n(x) \tag{3.40}$$

and

$$U_n=a_n(M_n(x)-Y_n)\,. \tag{3.41}$$

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

$$\lim_{n\to\infty}\theta_n=\theta\,. \tag{3.42}$$

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

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

$$\varepsilon_n=\Bigl(1+\frac{C_2}{C_1}\Bigr)|\theta-\theta_n|+\pi_n \text{ where } \{\pi_n\} \text{ is defined in the proof of Corollary 3.1.1 for } n\geqq 1\,, \tag{3.43}$$

and noting $|x-\theta|\leqq|\theta-\theta_n|+|x-\theta_n|$, then we have $|x-\theta_n|>\frac{C_2}{C_1}|\theta-\theta_n|+\pi_n$ for

$|x-\theta|>\varepsilon_n$. Hence, we have

$$(x-\theta)M_n(x)\geqq C_1\pi_n^2>0 \qquad \text{for} \quad |x-\theta|>\varepsilon_n \tag{3.44}$$

and

$$\inf_{\varepsilon_n<|x-\theta|<\varepsilon_n^{-1}} |M_n(x)|>C_1\pi_n \qquad \text{for all } n\geqq 1. \tag{3.45}$$

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

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 θ be the root of the equation $M(x)=0$. And let $\{\theta_n\}$ be a sequence of the roots of the equations $M_n(x)=0$, $n=1,2,\cdots$. Suppase 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 $n\geqq 1$.*

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

$$\mathrm{E}|X_n-\theta|^2\leqq Cv_n^\lambda \qquad \textit{for all } n\geqq 1. \tag{3.46}$$

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

$$\mathrm{P}[|X_n-\theta|^2\leqq C(\delta)v_n^\lambda \textit{ for all } n\geqq 1]>1-\delta. \tag{3.47}$$

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

$$\begin{aligned} &\mathrm{E}[|X_{n+1}-\theta|^2|X_1,\cdots,X_n] \\ &\leqq|X_n-\theta|^2+a_n^2K_n+C_2^2a_n^2|X_n-\theta_n|^2 \\ &\quad +2C_2a_n|X_n-\theta_n||\theta_n-\theta|-2C_1a_n|X_n-\theta_n|^2 \qquad a.s. \qquad \text{for all } n\geqq 1. \end{aligned} \tag{3.48}$$

Noting $X_n-\theta_n=X_n-\theta+\theta-\theta_n$ and $C_2\geqq C_1$, we have

$$C_2^2a_n^2|X_n-\theta_n|^2\leqq\frac{k+1}{k}C_2^2a_n^2|X_n-\theta|^2+(k+1)C_2^2a_n^2|\theta_n-\theta|^2 \tag{3.49}$$

and

$$\begin{aligned} &2C_2a_n|X_n-\theta_n||\theta_n-\theta|-2C_1a_n|X_n-\theta_n|^2 \\ &\leqq 6C_2a_n|X_n-\theta||\theta-\theta_n|+2C_2a_n|\theta-\theta_n|^2-2C_1a_n|X_n-\theta|^2 \\ &\leqq\frac{a_nC_1}{k}|X_n-\theta|^2+\frac{9C_2a_n}{C_1}|\theta_n-\theta|^2+2C_2a_n|\theta-\theta_n|^2-2C_1a_n|X_n-\theta|^2 \end{aligned} \tag{3.50}$$

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

$$\begin{aligned} &\mathrm{E}[|X_{n+1}-\theta|^2|X_1,\cdots,X_n] \\ &\leqq\Big[1-a_nC_1-\Big\{C_1a_n\Big(1-\frac{1}{k}\Big)-\frac{k+1}{k}C_2^2a_n^2\Big\}\Big]|X_n-\theta|^2 \\ &\qquad +\hat{C}(k)(a_n|\theta_n-\theta|+a_n^2K_n) \qquad a.s. \qquad \text{for all } n\geqq 1. \end{aligned} \tag{3.51}$$

where $\hat{C}(k)$ 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 $\frac{k_0-1}{k_0+1}C_1/C_2^2>a_n$ for all $n\geqq 1$. By virtue of (2.28) such a number k_0 exists. Hence, from (3.51), we have

$$\begin{aligned}(3.52)\qquad &\mathrm{E}[|X_{n+1}-\theta|^2|X_1,\cdots,X_n]\\ &\leqq(1-a_nC_1)|X_n-\theta|^2+\hat{C}(k_0)(a_n|\theta_n-\theta|+a_n^2K_n)\qquad a.s.\qquad \text{for all } n\geqq 1,\end{aligned}$$

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

§ 4. *H*-case.

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 $\mathrm{E}[Y_n(x)]=\mathbf{0}$ for sufficiently large n, where $Y_n(x)$ is an observed random element in H at $x\in H$ and at time n and $\mathbf{0}$ 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 $X_2, X_3,\cdots$ in accordance with the recurrence relation:

$$(4.1)\qquad X_{n+1}=X_n-a_nY_n,$$

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 $\mathscr{H}$-measurable transformation of H into itself, denoted by

$$(4.2)\qquad \mathrm{E}[Y_n(x)]=M_n(x),$$

where the expectation operator E is defined § 2.

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

$$(4.3)\qquad \mathrm{E}[Y_n|X_1,\cdots,X_n]=\mathrm{E}[Y_n(X_n)|X_n]=M_n(X_n)\qquad a.s..$$

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

$$(4.4)\qquad \sum_{n=1}^{\infty}a_n=+\infty,$$

$$(4.5)\qquad \sum_{n=1}^{\infty}a_n^2<+\infty.$$

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 $M_n(x)=\mathbf{0}$, $n=1,2,\cdots$. 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* $\|x-\theta_n\|^2>\varepsilon_n$ *for* $n\geqq 1$,

(ii) $\inf_{\varepsilon_n<\|x-\theta_n\|^2<\varepsilon_n^{-1}} |(x-\theta_n, M_n(x))|\geqq \rho_n$ *for* $n\geqq 1$,

(iii) $\|M_n(x)\|^2\leqq A_n(1+\|x-\theta_n\|^2)$ *for all* $x\in H$ *and* $n\geqq 1$,

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

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

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

Then it holds that

$$\lim_{n\to\infty} \mathrm{E}\|X_n-\theta_n\|^2=0\,, \tag{4.6}$$

and

$$\lim_{n\to\infty} \|X_n-\theta_n\|=0 \qquad a.\,s. \tag{4.7}$$

REMARK. (i) θ_n may not be the unique root of the equation $M_n(x)=\mathbf{0}$.

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

PROOF. Without loss of generality we can assume that

$$0<a_nA_n\leqq 1\,,\quad 0<a_n\leqq 1 \qquad \text{for all } n\geqq 1\,. \tag{4.8}$$

Define the measurable transformation T_n by

$$T_n(h)=h+\theta_n-\theta_{n+1}-a_nM(h+\theta_n) \qquad \text{for} \quad h\in H\,. \tag{4.9}$$

And we put

$$U_n=a_n(M(X_n)-Y_n)\,. \tag{4.10}$$

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

$$X_{n+1}-\theta_{n+1}=T_n(X_n-\theta_n)+U_n \qquad \text{for all } n\geqq 1\,. \tag{4.11}$$

Putting $Z_n=X_n-\theta_n$, we have

$$Z_{n+1}=T_n(Z_n)+U_n \qquad \text{for all } n\geqq 1\,. \tag{4.12}$$

And from (iv) and Assumption II we have

$$\sum_{n=1}^{\infty} \mathrm{E}\|U_n\|^2<\infty \tag{4.13}$$

and

$$\mathrm{E}[U_n|Z_1, Z_2, \cdots, Z_n]=\mathbf{0} \qquad a.\,s. \qquad \text{for all } n\geqq 1\,. \tag{4.14}$$

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

$$\|T_n(h)\|^2 \leqq \|h\|^2+\|\theta_n-\theta_{n+1}\|^2+a_n^2\|M_n(h+\theta_n)\|^2+2\|h\|\|\theta_n-\theta_{n+1}\| \\ +2a_n\|\theta_n-\theta_{n+1}\|\|M_n(h+\theta_n)\|-2a_n(h, M(h+\theta_n)) \quad \text{for all } n\geqq 1. \tag{4.15}$$

Noting

$$2\|h\|\cdot\|\theta_n-\theta_{n+1}\|\leqq a_n^2\|h\|^2+\frac{1}{a_n^2}\|\theta_n-\theta_{n+1}\|^2$$

and

$$2a_n\|\theta_n-\theta_{n+1}\|\|M_n(h+\theta_n)\|\leqq\|\theta_n-\theta_{n+1}\|^2+a_n^2\|M_n(h+\theta_n)\|^2,$$

from (iii) we have

$$\|T_n(h)\|^2\leqq(1+a_n^2+2a_n^2A_n)\|h\|^2 \\ +2a_n^2A_n+\Bigl(2+\frac{1}{a_n^2}\Bigr)\|\theta_n-\theta_{n+1}\|^2-2a_n(h, M(h+\theta_n)) \\ \text{for all } h\in H \text{ and } n\geqq 1. \tag{4.16}$$

Suppose $\varepsilon_n<\|h\|^2<\varepsilon_n^{-1}$. From (ii) and (4.16) we have

$$\|T_n(h)\|^2\leqq(1+a_n^2+2a_n^2A_n)\|h\|^2+2a_n^2A_n+\frac{(1+2a_n^2)}{a_n^2}\|\theta_n-\theta_{n+1}\|^2-2a_n\rho_n \\ \text{for all } n\geqq 1. \tag{4.17}$$

Suppose $\varepsilon_n\geqq\|h\|^2$. Noting $2a_n|(h, M_n(h+\theta_n))|\leqq a_n\{\|h\|^2+A_n(1+\|h\|^2)\}$, we have

$$\|T_n(h)\|^2\leqq(1+a_n+a_nA_n+a_n^2+2a_n^2A_n)\varepsilon_n+2a_n^2A+\frac{(1+2a_n^2)}{a_n^2}\|\theta_n-\theta_{n+1}\|^2 \\ \text{for all } n\geqq 1. \tag{4.18}$$

Next, suppose that $\varepsilon_n^{-1}\leqq\|h\|^2$, then we have

$$\|T_n(h)\|^2\leqq(1+a_n^2+2a_n^2A_n)\|h\|^2+2a_n^2A_n+\frac{(1+2a_n^2)}{a_n^2}\|\theta_n-\theta_{n+1}\|^2 \\ \text{for all } n\geqq 1. \tag{4.19}$$

Hence, from (4.17), (4.18), (4.19) and (4.8), we have

$$\|T_n(h)\|^2\leqq\max\Bigl\{6\varepsilon_n+2a_n^2A_n+\frac{3}{a_n^2}\|\theta_n-\theta_{n+1}\|^2, (1+a_n^2+2a_n^2A_n)\|h\|^2 \\ -\gamma_n(h)+2a_n^2A_n+\frac{3}{a_n^2}\|\theta_n-\theta_{n+1}\|^2\Bigr\} \tag{4.20}$$

where

$$\begin{aligned}\gamma_n(h)&=-2a_n\rho_n &&\text{if } \|h\|^2<\varepsilon_n^{-1}\\ &=0 &&\text{if } \|h\|^2\geqq\varepsilon_n^{-1}.\end{aligned} \tag{4.21}$$

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 θ_n is the unique root of the equation $M_n(x)=0$.

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

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

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

$$C_1\|x-\theta_n\|^2\leqq(M_n(x), x-\theta_n),$$
$$\|M_n(x)\|\leqq C_2\|x-\theta_n\| \quad \textit{for all } x\in H \textit{ and } n\geqq 1,$$

(iii) $\sum_{n=1}^{\infty} a_n \mathrm{E}\|M_n(X_n)-Y_n\|^2<\infty$,

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 $\varepsilon_n=\pi_n$, where $\{\pi_n\}$ 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

$$(M_n(x),\, x-\theta_n)\geqq C_1\|x-\theta_n\|^2>C_1\varepsilon_n>0\,.$$

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 $M_n(x)=\mathbf{0}$, $n\geqq 1, 2, \cdots$. 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*

$$\mathrm{E}[\|M_n(X_n)-Y_n\|^2\,|\,X,\cdots,X_n]\leqq K_n \qquad a.s. \qquad \textit{for all } n\geqq 1\,. \tag{4.22}$$

(1) *If there exists a number $0<\lambda<1$ satisfying* (3.25) *for all $n\geqq N$ and* (3.26) *for $v_n=\max\left\{\frac{\|\theta_n-\theta_{n+1}\|^2}{a_n},\, a_n^2K_n\right\}$ then there exists a constant $C>0$ such that*

$$\mathrm{E}\|X_n-\theta_n\|^2\leqq Cv_n^\lambda \qquad \textit{for all } n\geqq 1\,. \tag{4.23}$$

(2) *If there exists a number $0<\lambda<1$ satisfying* (3.25) *for all $n\geqq 1$ and* (3.26) *with $v_n=\max\left\{\frac{\|\theta_n-\theta_{n+1}\|^2}{a_n},\, a_n^2K_n\right\}$, and if* (3.28) *is satisfied, then for any $\delta>0$ there exists a constant $C(\delta)>0$ such that*

$$\mathrm{P}[\|X_n-\theta_n\|^2\leqq C(\delta)v_n^\lambda \textit{ for all } n\geqq 1]>1-\delta\,. \tag{4.24}$$

PROOF. From (4.1) and Assumption II we have

$$\begin{aligned}
&\mathrm{E}[\|X_{n+1}-\theta_{n+1}\|^2\,|\,X_1,\cdots,X_n] \\
&\quad=\|X_n-\theta_n\|^2+\|\theta_n-\theta_{n+1}\|^2+a_n^2\,\mathrm{E}[\|M_n(X_n)-Y_n\|^2\,|\,X_1,\cdots,X_n] \\
&\qquad+a_n^2\,\|M_n(X_n)\|^2-2a_n(X_n-\theta_n,\,M_n(X_n))+2(X_n-\theta_n,\,\theta_n-\theta_{n+1}) \\
&\qquad+2a_n(M_n(X_n),\,\theta_n-\theta_{n+1}) \qquad a.s. \qquad \text{for all } n\geqq 1\,.
\end{aligned} \tag{4.25}$$

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 $M(x)=\mathbf{0}$ which is independent of the time n.

Let us assume that in place of a random element $Y(x)$ in H satisfying $\mathrm{E}[Y(x)]=M(x)$, 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 $\mathrm{E}[Y_n(x)]$ (say $M_n(x)$ coincides with $M(x)$, nor that the equation $M_n(x)=\mathbf{0}$ has a root. Under these situations, however, we shall follow the procedure (4.1).

THEOREM 4.3. *Let θ be the root of the equation $M(x)=\mathbf{0}$. Suppose that there exist three sequences $\{\varepsilon_n\}$, $\{\rho_n\}$, $\{A_n\}$ satisfying*

(i) $(x-\theta, M_n(x))>0$ *if* $\|x-\theta\|^2>\varepsilon_n$ *for all* $n\geqq 1$,

(ii) $\inf_{\varepsilon<\|x-\theta\|^2<\varepsilon_n^{-1}} (x-\theta, M_n(x))\geqq \rho_n$ *for all* $n\geqq 1$,

(iii) $\|M_n(x)\|^2\leqq A_n(1+\|x-\theta\|^2)$ *for all* $x\in H$ *and* $n\geqq 1$,

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

and

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

Then it holds that

$$\lim_{n\to\infty} \mathrm{E}\|X_n-\theta\|^2=0 \tag{4.26}$$

and

$$\lim_{n\to\infty} \|X_n-\theta\|=0 \qquad a.\ s. \tag{4.27}$$

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

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

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

$$\lim_{n\to\infty} \|\theta_n-\theta\|^2=0\,. \tag{4.28}$$

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

PROOF. Putting $\varepsilon_n=\left(2+\frac{2C_2}{C_1}\right)\|\theta_n-\theta\|^2+\pi_\pi^{\frac{1}{2}}$, where $\{\pi_n\}$ was defined in the proof of Corollary 3.1.1, and $A_n=\max\{2C_2\|\theta_n-\theta\|^2, 2C_2\}$, we get (4.30) and (4.31) by Theorem 4.1. And noting $\|X_n-\theta_n\|^2\leqq 2\|X_n-\theta\|^2+2\|\theta_n-\theta\|^2$, 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 $M_n(x)=\mathbf{0}$, $n=1, 2, \cdots$. Let θ be the root of the equation $M_n(x)=\mathbf{0}$. 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 satisfie* (4.22) *for all $n\geqq 1$.*

(1) *If there exists a number $0<\lambda<1$ which satisfies* (3.25) *for all $n\geqq N$ 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*

$$\mathrm{E}\|X_n-\theta\|^2\leqq Cv_n^\lambda \qquad \text{for all } n\geqq 1\,. \tag{4.29}$$

(2) *If there exists a number $0<\lambda<1$ which satisfies* (3.25) *for all $n\geqq 1$ 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*

$$\mathrm{P}[\|X_n-\theta\|^2\leqq C(\delta)v_n^\lambda \text{ for all } n\geqq 1]>1-\delta\,. \tag{4.30}$$

PROOF. From (4.1) and Assumption II we have

$$\begin{aligned}(4.31) \quad & \mathrm{E}[\|X_{n+1}-\theta\|^2 | X_1, \cdots, X_n] \\ & = \|X_n-\theta\|^2 - 2a_n(X_n-\theta, M_n(X_n)) \\ & \quad + a_n^2 \|M_n(X_n)\|^2 + a_n^2 \mathrm{E}[\|M_n(X_n)-Y_n\|^2 | X_1, \cdots, X_n] \quad a.s. \quad \text{for all } n \geqq 1.\end{aligned}$$

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 $\hat{M}(x)$ 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 $\hat{M}(x)=0$. Let $\{\hat{Y}_n(x)\}$ be a sequence of random variable with parameter $x \in H$. We put

$$\mathrm{E}\hat{Y}_n(x) = \hat{M}_n(x) \quad \text{for } n \geqq 1 \text{ and } x \in H. \tag{5.1}$$

In this problem, we follow the procedure:

$$\begin{aligned} & X_1 \text{ is an arbitrary fixed element in } H, \\ & X_{n+1} = X_n - a_n \hat{Y}_n e \quad \text{for } n \geqq 1, \end{aligned} \tag{5.2}$$

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 $Y_n(x)=\hat{Y}_n(x)e$ and $Y_n=\hat{Y}_n e$, 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 $\hat{f}$ w. r. p. the Lebesgne measure on R. Let us assume $\hat{f} \in L^2(R)$ 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

$$\sup_y K(y) < \infty, \tag{5.3}$$

$$\int_R K(y)dy = 1, \tag{5.4}$$

$$\lim_{y \to \infty} |y| K(y) = 0. \tag{5.5}$$

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

$$K_n(y, z) = \frac{1}{h_n} K\left(\frac{y-z}{h_n}\right), \tag{5.6}$$

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

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

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

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

$$\begin{aligned} & f_0 \equiv 0, \\ & f_{n+1}(y) = f_n(y) - a_n\{f_n(y) - K_n(y, Z_{n+1})\} \quad \text{for } n \geqq 1 \text{ and } y \in R, \end{aligned} \tag{5.8}$$

where $\{a_n\}$ is a sequence satisfying (4.4) and (4.5). In § 4, we put $H=L^2(R)$, $X_n=f_n$, $Y_n=f_n-K_n(\cdot, Z_{n+1})$, $Y_n(f)=f-K_n(\cdot, Z_{n+1})$, $\theta=\hat{f}$ and $\theta_n=\hat{f}_n$, then we have

$$\mathrm{E}[Y_n(f)]=f-\hat{f}_n \qquad \text{for any } f\in L^2(R)\,, \tag{5.9}$$

$$\begin{aligned}\mathrm{E}[Y_n|X_1, \cdots, X_n]&=\mathrm{E}[f_n-K_n(\cdot, Z_{n+1})|f_1, \cdots, f_n]\\ &=f_n-\hat{f}_n \qquad a.s.\end{aligned} \tag{5.10}$$

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 $\lim_{n\to\infty}\mathrm{E}\|f_n-\hat{f}\|^2=0$.

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 $\lim_{n\to\infty}\mathrm{E}|f_n(y)-\hat{f}(y)|=0$ for all y and that $\lim_{n\to\infty}|f_n(y)-f(y)|=0$ $a.s.$ for all y ([7]).

(3) *The learning problem for pattern classification* ([9], [13], [14], [15]). Let us consider the problem of estimating a unknown discriminent function $\hat{D}(y)=q_A f_B(y)-q_B f_B(y)$, 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 ([13], [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):

$$\begin{aligned}&D_0\equiv 0\,,\\ &D_{n+1}(y)=D_n(y)-a_{n+1}\{D_{n+1}(y)-\rho(\Theta_{n+1})K_n(y, Z_{n+1})\} \qquad \text{for each } y\in R\,,\end{aligned} \tag{5.11}$$

where

$$\begin{aligned}\rho(\Theta_n)&=1 \qquad \text{if } \Theta_n=A\\ &=-1 \qquad \text{if } \Theta_n=B\,.\end{aligned}$$

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 $\hat{D}_n(y)=\mathrm{E}[\rho(\Theta_{n+1})K_n(y, Z_{n+1})]$ for $y\in R$. Then we also have

$$\mathrm{E}[Y_n(D)]=D-D_n \qquad \text{for each } D\in L^2(R) \text{ and } n\geqq 1\,, \tag{5.12}$$

$$\begin{aligned}\mathrm{E}[Y_n|X_1, \cdots, X_n]&=\mathrm{E}[D_n-\rho(\Theta_{n+1})K_n(\cdot, Z_{n+1})|D_1, \cdots, D_n]\\ &=D_n-\hat{D}_n \qquad a.s. \qquad \text{for } n\geqq 1\,.\end{aligned} \tag{5.13}$$

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 indepted to Professor N. Furukawa of Kyushu University for helpful advices and encouragements.

References

[1] E.M. BRAVERMAN and E.S. PYATNITSKII, *Estimation of the rate of convergence algorithms base on the potential function method.* Autom. and Remote Cont., **27**, No. 1 (1966), 80–100.

[2] J.R. BLUM, *Multidimensional stochastic approximation methods.* Ann. Math. Stat., **25** (1954), 737–744.

[3] V. DUPAC, *A dynamic stochastic approximation method.* Ann. Math. Stat., **35** (1965), 1695–1702.

[4] V. DUPAC and F. KRAL, *Robbins-Monro procedure with both variables subject to experimental error.* Ann. Math. Stat., **43**. No. 4, (1972), 1089–1095.

[5] A. DVORETZKY, *On stochastic approximation.* Proc. Third Berkeley Symp. Math. Stat. Prob., **1** (1956), 39–55.

[6] N.V. LOGINOV, *The method of stochastic approximation.* Autom. and Remote. Cont., **27**, No. 4, (1966), 706–728.

[7] E. PARZEN, *On the estimation of probability density function and mode.* Ann. Math. Stat., **33** (1962), 1065–1976.

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

[9] K. TANAKA, *On the pattern classification problem by learning.* Bull. Math. Stat., **14**, No. 3–4, (1971), 15–25.

[10] J.H. VENTER, *On Dvoretzky stochastic approximation method.* Ann. Math. Stat., **37** (1966), 1534–1544.

[11] H. YAMATO, *Sequential estimation of a continuous probability density function and mode.* Bull. Math. Stat., **14**, No. 3–4, (1971), 1–12.

[12] T.Y. YOUNG and R.A. WESTERBERG, *Stochastic approximation with a non-stationary regression function.* IEEE Trans. on Information Theory, **IT-18**, No. 4, (1972), 518–519.

[13] M. WATANABE, *On asymptotically optimal algorithms for pattern classification problems.* Bull. Math. Stat., **15**, No. 3–4, (1973), 31–48.

[14] M. WATANABE, *On convergences of asymptotically optimal discriminant functions for a pattern classification problem.* Bull. Math. Stat., **16**, No. 1–2, (1974), 23–34.

[15] M. WATANABE, *On a pattern classification problem on the basis of a training sequence associated with dependent random variables.* Bull. Math. Stat., **16**, No. 1–2, (1974), 35–48.