Z. Wahrscheinlichkeitstheorie verw. Gebiete
54, 47–55 (1980)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Stochastic Approximation from Ergodic Sample for Linear Regression

László Györfi

Technical University of Budapest, Stoczek u. 2.,
H-1111 Budapest, Hungary

Summary. Robbins-Monro stochastic approximation procedure $x_{n+1}=x_n-\frac{1}{n+1}(A_{n+1}x_n-y_{n+1})$ is used to solve the linear equation $Ax=y$ in Hilbert space, where y_n and A_n are estimators such that their arithmetic means converge to y and A, respectively. Under some additional conditions it is shown that X_n goes to the unique solution of this equation.

Introduction: Linear Regression

Some problems of prediction, filtering, pattern classification, control and system identification can be formulated by the following linear regression: let ξ and η be N dimensional vector valued random variables and the question of interest is the solution of the linear equation

$$Ax=y \tag{1}$$

where $A=\mathbb{E}(\xi\xi^T)$ and $y=\mathbb{E}\eta$. (The vectors are column vectors, T stands for the transposition.) We are given a dependent sample $(\xi_1,\eta_1),(\xi_2,\eta_2)\dots$ where only the strong law of large numbers can be assumed, namely

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\xi_i\xi_i^T=A \quad \text{a.s.,} \tag{2}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\eta_i=y \quad \text{a.s.} \tag{3}$$

0044-3719/80/0054/0047/$01.80

Assume that A^{-1} exists, then for large n $\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\xi_i^T\right)^{-1}$ exists and

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\xi_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)=A^{-1}y \quad \text{a.s.} \tag{4}$$

However, from the point of view of application (4) is very complicated, therefore we are interested in Robbins-Monro stochastic approximation: x_0 is arbitrary,

$$x_{n+1}=x_n-\frac{1}{n+1}(\xi_{n+1}\xi_{n+1}^T x_n-\eta_{n+1}). \tag{5}$$

Proposition. *If, in addition,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\|\xi_i\|^4 \quad \textit{exists a.s.} \tag{6}$$

then for x_n defined by (5)

$$\lim_{n\to\infty} x_n=A^{-1}y \quad \textit{a.s.} \tag{7}$$

Observe that (5) does not use real matrix operation, since (5) might be written in form

$$x_{n+1}=x_n-\frac{1}{n+1}((\xi_{n+1},x_n)\,\xi_{n+1}-\eta_{n+1}), \tag{8}$$

where $(.,.)$ denotes the inner product in R^N.

Main Result

In the sequel we formulate a natural extension of this problem to Hilbert space as Venter (1966) and Révész (1973) made for more general stochastic approximation procedures.

Let H be a real Hilbert space with the inner product $(.,.)$ and norm $\|.\|$. Denote by A an unknown linear, bounded, symmetric and positive operator on H, and we have to solve the equation $Ax=y$ for an unknown $y\in H$. Assume that A^{-1} exists. We are given a sequence of linear, bounded operators $A_1, A_2, \ldots$ and a sequence $y_n\in H, n=1,2\ldots$.

Theorem 1. *Suppose that*

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}y_i-y\right\|=0, \tag{9}$$

and

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}A_i-A\right\|=0. \tag{10}$$

Assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|A_i\|^2 \tag{11}$$

exists. Consider the sequence x_n: x_0 *is arbitrary,*

$$x_{n+1} = x_n - \frac{1}{n+1}(A_{n+1} x_n - y_{n+1}) \tag{12}$$

then

$$\lim_{n\to\infty} x_n = A^{-1} y. \tag{13}$$

Fritz (1974) investigated the same problem for linear bounded operators in Banach space, under a contraction-type condition. Specializing to symmetric positive operators in Hilbert spaces, he obtained our Theorem 1, under the assumption $\|A_i\| \leqq 1$ $i=1,2\dots$ instead of our assumption (11). Csibi (1973) and (1975) showed a.s. convergence of general stochastic approximation for m_0-dependent and uniformly strong mixing sample. Ljung (1978) dealt with the recursion

$$x_{n+1} = x_n - \gamma_n (f(x_n) + y_{n+1})$$

where in case of $\gamma_n = \dfrac{1}{n+1}$ only (9) is required on the additive noise y_i's. If f is linear, then his result implies the a.s. convergence provided $A_i = A, i = 1, 2, \dots$

There are some accelerated versions of the iteration (8) (Tsypkin (1970) and Saridis, Nikolic, Fu (1969)). For dependent sample the convergence of their algorithms may be deduced from the following formal extension of Theorem 1:

Theorem 2. *Consider the iteration:*
x_0 *is arbitrary.*

$$x_{n+1} = x_n - \frac{c_1}{c_2 + n} U_{n+1}(A_{n+1} x_n - y_{n+1}), \tag{14}$$

where U_n $n=1,2\dots$ *are linear bounded operators such that* $\lim_{n\to\infty} \|U_n - U\| = 0$. A^{-1} *and* U^{-1} *exists.* UA *is symmetric and positive. Assume* (9). (10) *and* (11). *Then* $\lim x_n = A^{-1} y$. ■

Theorem 2 can be easily verified from Theorem 1 since for the notation $A'_{n+1} = \dfrac{c_1(n+1)}{c_2+n} U_{n+1} A_{n+1}$ and $y'_{n+1} = \dfrac{c_1(n+1)}{c_2+n} U_{n+1} y_{n+1}$ the conditions of Theorem 2 imply the conditions of Theorem 1 for the iteration

$$x_{n+1} = x_n - \frac{1}{n+1}(A'_{n+1} x_n - y'_{n+1}), \tag{15}$$

therefore x_n tends to the unique solution of $UAx = Uy$ which is the same as that of $Ax = y$.

Proofs. An abstract version of the well-known Toeplitz Theorem is applied several times during the proof of Theorem 1:

Toeplitz Theorem (see Fritz (1974)). Consider a triangular array $\mathscr{C}_{k,n}\, k=1,\dots n$, $n=1,2\dots$ of linear, bounded operators on a Banach space $\mathscr{B}$, for which

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathscr{C}_{k,n}x=\mathscr{C}x$$

for each $x\in\mathscr{B}$ and for each fixed integer k

$$\lim_{n\to\infty}\|\mathscr{C}_{k,n}\|=0.$$

If

$$\sup_{n}\sum_{k=1}^{n}\|\mathscr{C}_{k,n}\|<+\infty,$$

then $\lim_{n\to\infty}x_n=x$ implies that $\lim_{n\to\infty}\sum_{k=1}^{n}\mathscr{C}_{k,n}x_k=\mathscr{C}x$. If

$$\sup_{n}\left[(n+1)\|\mathscr{C}_{n,n}\|+\sum_{k=1}^{n-1}(k+1)\,\|\mathscr{C}_{k,n}-\mathscr{C}_{k+1,n}\|\right]<+\infty$$

then $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}x_k=x$ implies that $\lim_{n\to\infty}\sum_{k=1}^{n}\mathscr{C}_{k,n}x_k=\mathscr{C}x$.

Proof of Theorem 1. It is sufficient to deal with the case of $y=0$, since from (12)

$$x_{n+1}-A^{-1}y=x_n-A^{-1}y-\frac{1}{n+1}(A_{n+1}(x_n-A^{-1}y)-(y_{n+1}-A_{n+1}A^{-1}y))$$

and using the notations $x_n'=x_n-A^{-1}y$ and $y_{n+1}'=y_{n+1}-A_{n+1}A^{-1}y$ we get the same iteration as that of (12) and (9), (10) imply

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}y_i'=0,$$

therefore in the sequel we suppose that $y=0$. Put

$$B_{k,n}=\begin{cases}\frac{1}{k}I & \text{if } k=n\\ \frac{1}{k}\left(I-\frac{1}{n}A_n\right)\left(I-\frac{1}{n-1}A_{n-1}\right)\dots\left(I-\frac{1}{k+1}A_{k+1}\right) & \text{if } 0<k<n\\ \left(I-\frac{1}{n}A_n\right)\qquad\dots(I-A_1) & \text{if } k=0\end{cases}\tag{16}$$

and $y_0=x_0$, then we get from (12) by induction that

$$x_n=\sum_{k=0}^{n}B_{k,n}y_k\qquad n=1,2\dots,\tag{17}$$

and for the notation $s_n=\frac{1}{n}\sum_{k=1}^{n} y_k$ $n=1,2\ldots$, $s_0=y_0=x_0$

$$x_n=\sum_{k=0}^{n} B'_{k,n} s_k \tag{18}$$

where

$$B'_{k,n}=\begin{cases} I & \text{if } k=n \\ B_{k+1,n}(I-A_{k+1}) & \text{if } 0<k<n \\ B_{0,n} & \text{if } k=0 \end{cases} \tag{19}$$

From (18)

$$\|x_n\| \leqq \sum_{k=0}^{n} \|B'_{k,n}\| \, \|s_k\| \tag{20}$$

By (9) $\lim_{n\to\infty} s_n=y=0$; therefore for (20) we could apply an other version of Toeplitz Theorem (see Ash (1972) 7.1.1. Lemma) if we knew that for each fixed integer k

$$\lim_{n\to\infty} \|B'_{k,n}\|=0 \tag{21}$$

and

$$\sup_n \sum_{k=0}^{n} \|B'_{k,n}\|<+\infty. \tag{22}$$

(20), (21) and (22) imply $\lim_{n\to\infty} x_n=0$. To prove (21) and (22) it will be useful the following

Lemma. *Put* $m=\inf_{\|u\|=1}(Au,u)$. *(*A *is positive and* A^{-1} *exists, therefore* $m>0$.*) Under the conditions* (10) *and* (11) *for each* $\delta<m$ *there exists a real* $C>1$ *such that for each* $k\geqq n$

$$\|B_{k,n}\| \leqq \begin{cases} c\dfrac{1}{k}\left(\dfrac{k}{n}\right)^{\delta} & \text{if } k\geqq 1 \\ c\left(\dfrac{1}{n}\right)^{\delta} & \text{if } k=0 \end{cases} \tag{23}$$

The proof of Lemma will be given later.

Continuing the proof of Theorem 1 (19) and (23) imply (21). By (19)

$$\sum_{k=0}^{n-1} \|B'_{k,n}\| \leqq \sum_{k=1}^{n} \|B_{k,n}\| \,(1+\|A_k\|) \tag{24}$$

Let $r_n=\frac{1}{n}\sum_{k=1}^{n}(1+\|A_k\|)$, then (11) implies that

$$\limsup_{n\to\infty} r_n=L^*<+\infty \tag{25}$$

and by (23), (24)

$$\begin{aligned}\sum_{k=0}^{n-1} \|B'_{k,n}\| &\leqq C \sum_{k=1}^{n} \frac{1}{k}\left(\frac{k}{n}\right)^{\delta} (1+\|A_k\|)\\ &= C \sum_{k=1}^{n} \frac{1}{k}\left(\frac{k}{n}\right)^{\delta} (k r_k-(k-1) r_{k-1})\\ &= C\left[\sum_{k=1}^{n-1} \left(\frac{1}{k}\left(\frac{k}{n}\right)^{\delta}-\frac{1}{k+1}\left(\frac{k+1}{n}\right)^{\delta}\right) k r_k + r_n\right]. \end{aligned} \tag{26}$$

If $\delta>1$ then from (25) and (26)

$$\limsup_{n\to\infty} \sum_{k=0}^{n} \|B'_{k,n}\| \leqq C L^* + 1$$

otherwise

$$\limsup_{n\to\infty} \sum_{k=0}^{n} \|B'_{k,n}\| \leqq C \frac{L^*}{\delta} + 1.$$

Thus the proof of Theorem 1 is complete.

Proof of Lemma. First we prove that there exists an integer N_0 and a real $C'>1$ such that for each $N_0 \leqq k \leqq n$

$$\|B_{k,n}\| \leqq C' \frac{1}{k}\left(\frac{k}{n}\right)^{\delta} \tag{27}$$

We use the following version of induction: for $n=k$

$$\|B_{k,k}\| = \left\|\frac{1}{k} I\right\| = \frac{1}{k} \leqq \frac{C'}{k}. \tag{28}$$

Assume that for each i, $k \leqq i \leqq n$

$$\|B_{k,i}\| \leqq C' \frac{1}{k}\left(\frac{k}{i}\right)^{\delta} \tag{29}$$

then we show that for sufficiently large C' there exists an integer N_0 such that for each $n \geqq k \geqq N_0$ (29) imply that

$$\|B_{k,n+1}\| \leqq C' \frac{1}{k}\left(\frac{k}{n+1}\right)^{\delta}. \tag{30}$$

Let us denote by L the $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|A_i\|^2$. Choose $L'>m\delta$ such that

$$2m - \frac{m}{L'} L > 2\delta - \frac{m\delta^2}{L'}. \tag{31}$$

Introduce the notations

$$\alpha_i = \left(\frac{i}{n+1}\right)^{\frac{L'}{m}} \tag{32}$$

and

$$\beta_i = \frac{(i+1)^{\frac{L'}{m}} - i^{\frac{L'}{m}}}{i^\delta} \tag{33}$$

and

$$g_i = \frac{L'}{m} \frac{(i+1)^{\frac{L'}{m}}}{(i+1)\left((i+1)^{\frac{L'}{m}} - i^{\frac{L'}{m}}\right)} \tag{34}$$

and

$$S_n = \sum_{i=1}^{n} \beta_i. \tag{35}$$

Then $\lim_{n\to\infty} g_n = 1$ and by (16)

$$\begin{aligned} B_{k,n+1} &= \sum_{i=k}^{n} (\alpha_{i+1} B_{k,i+1} - \alpha_i B_{k,i}) + \alpha_k B_{k,k} \\ &= \sum_{i=k}^{n} (\alpha_{i+1} - \alpha_i) \left(I - \frac{m}{L'} g_i A_{i+1}\right) B_{k,i} + \frac{\alpha_k}{k} I. \end{aligned} \tag{36}$$

Denote B^* the adjoint of the operator B, then by (36)

$$\begin{aligned} \|B_{k,n+1}\| &= \|B^*_{k,n+1}\| \\ &= \left\| \sum_{i=k}^{n} (\alpha_{i+1} - \alpha_i) B^*_{k,i} \left(I - \frac{m}{L'} g_i A^*_{i+1}\right) + \frac{\alpha_k}{k} I \right\| \\ &= \sup_{\|u\|=1} \left\| \left[\sum_{i=k}^{n} (\alpha_{i+1} - \alpha_i) B^*_{k,i} \left(I - \frac{m}{L'} g_i A^*_{i+1}\right) + \frac{\alpha_k}{k} I \right] u \right\| \\ &\leqq \sup_{\|u\|=1} \sum_{i=k}^{n} (\alpha_{i+1} - \alpha_i) \|B^*_{k,i}\| \cdot \left\| \left(I - \frac{m}{L'} g_i A^*_{i+1}\right) u \right\| + \frac{\alpha_k}{k}. \end{aligned} \tag{37}$$

Put

$$Z^u_n = \frac{1}{S_n} \sum_{i=1}^{n} \beta_i \left\| \left(I - \frac{m}{L'} g_i A^*_{i+1}\right) u \right\|^2 \tag{38}$$

then applying (29), (37) and the Cauchy-Schwarz inequality we get

$$\|B_{k,n+1}\| \leqq C' \frac{1}{k} \frac{k^\delta}{(n+1)^{\frac{L'}{m}}} S_n \left(\sup_{\|u\|=1} Z^u_n \right)^{\frac{1}{2}} + \frac{\alpha_k}{k}$$

$$\leq C' \frac{1}{k} \left(\frac{k}{n+1}\right)^{\delta} \left[\frac{S_n (\sup\limits_{\|u\|=1} Z_n^u)^{\frac{1}{2}}}{(n+1)^{\frac{L'}{m}-\delta}} + \frac{1}{C'}\right] \tag{39}$$

Observe that

$$\frac{S_n}{(n+1)^{\frac{L'}{m}-\delta}} = \delta'_n \frac{1}{1-\frac{m\delta}{L'}} \tag{40}$$

where $\lim\limits_{n\to\infty} \delta'_n = 1$. If we show that

$$\limsup_{n\to\infty} \sup_{\|u\|=1} Z_n^u \leq 1 - 2\frac{m^2}{L'} + \frac{m^2}{L'^2} L \tag{41}$$

then from (39), (40) and (41) we get

$$\|B_{k,n+1}\| \leq C' \frac{1}{k} \left(\frac{k}{n+1}\right)^{\delta} \left[\frac{\left(1 - 2\frac{m^2}{L'} + \frac{m^2}{L'^2} L + \delta''_n\right)^{\frac{1}{2}}}{1-\frac{m\delta}{L'}} \delta'_n + \frac{1}{C'}\right], \tag{42}$$

where $\lim\limits_{n\to\infty} \delta''_n = 0$, therefore because of (31) there exists an integer N_0 such that

$$\beta = \sup_{N_0 \leq n} \delta'_n \frac{\left(1 - 2\frac{m^2}{L'} + \frac{m^2}{L'^2} L + \delta''_n\right)^{\frac{1}{2}}}{1-\frac{m\delta}{L'}} < 1 \tag{43}$$

and for

$$C' > \frac{1}{1-\beta}$$

from (42)

$$\|B_{k,n+1}\| \leq C' \frac{1}{k} \left(\frac{k}{n+1}\right)^{\delta}.$$

In order to prove (41) we get from (38) that

$$\begin{aligned}
Z_n^u &= \|u\|^2 - 2\frac{m}{L'} \left(\frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i A_{i+1}^* u, u\right) \\
&\quad + \frac{m^2}{L'^2} \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i^2 \|A_{i+1}^* u\|^2 \\
&= \|u\|^2 - 2\frac{m}{L'} (A^* u, u) + 2\frac{m}{L'} \left(\left(A^* - \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i A_{i+1}^*\right) u, u\right) \\
&\quad + \frac{m^2}{L'^2} \left(\frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i^2 A_{i+1} A_{i+1}^* u, u\right)
\end{aligned}$$

$$\leqq \|u\|^2 \left(1 - 2\frac{m^2}{L'} + 2\frac{m}{L'}\left\| A - \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i A_{i+1} \right\| + \frac{m}{L'^2}\left\| \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i^2 A_{i+1} A_{i+1}^* \right\| \right). \tag{44}$$

Because of Toeplitz Theorem (10) implies that

$$\lim_{n\to\infty} \left\| A - \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i A_{i+1} \right\| = 0. \tag{45}$$

Applying (11) and the Toeplitz Theorem we get

$$\begin{aligned} &\limsup_{n\to\infty} \left\| \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i^2 A_{i+1} A_{i+1}^* \right\| \\ &\leqq \lim_{n\to\infty} \frac{1}{S_n} \sum_{i=1}^{n} \beta_i g_i^2 \|A_{i+1}\|^2 \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|A_i\|^2 = L \end{aligned} \tag{46}$$

(44), (45) and (46) imply (41), therefore the proof of (27) is complete. If $k \leqq N_0 \leqq n$, then by (16)

$$B_{k,n} = (N_0 + 1)\, B_{N_0+1,n} B_{k,N_0} \tag{47}$$

(16), (27) and (47) imply the statement of the Lemma.

References

Ash, R.B.: Real Analysis and Probability. New York: Academic Press 1972

Csibi, S.: Statistical learning processes. Preprint. Technical University of Budapest 1973

Csibi, S.: Learning under computational constraints from weakly dependent samples. Problems of Control and Information Theory. **4**, 3–21 (1975)

Fritz, J.: Learning from an ergodic training sequence. In Limit Theorems of Probability Theory; ed. P. Révész. Amsterdam: North-Holland 79–91 (1974)

Ljung, L.: Strong convergence of a stochastic approximation algorithm. Ann. Statist. **6**, 680–696 (1978)

Révész, P.: Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. Studia Sci. Math. Hungar. **8**, 391–398 (1973)

Saridis, G.N., Nikolic, Z.J., Fu, K.S.: Stochastic approximation algorithms for system indentification, estimation, and decomposition of mixtures. IEEE Trans. Systems Science and Cybernetics **5**, 8–15 (1969)

Tsypkin, Y.A.: Foundations of the Theory of Learning Systems. In Russian. Moscow: Nauka 1970

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

Received May 25, 1979