A RANDOM OBSERVATION PROCESS FOR STOCHASTIC APPROXIMATION

SIRŌ YAMAZOE

(Received June 23, 1970; revised April 3, 1972)

1. Introduction and summary

Consider a family of probability distributions $\{F_x\}$ with a controllable parameter (level) x and finite means $M(x)$. As is well known, the stochastic approximation method may be used to find the location x_0 at which the function M has a certain property. In fact Robbins and Monro [6] proposed the following procedure to estimate the root x_0 of the equation $M(x)=a$, assuming that M is an increasing function:

$$x_{n+1}=x_n+a_n(a-y_n) , \tag{1}$$

where $\{a_n\}$ is a fixed sequence of positive constants and y_n is a random variable with the conditional distribution F_{x_n}, given $x_1,\cdots,x_n,y_1,\cdots,y_{n-1}$. It is proved that x_n converges to x_0 with probability one under some conditions; see Blum [1], Dvoretzky [3] and Schmetterer [7]. In case $M(x)$ is strictly increasing for $x<x_0$ and decreasing for $x>x_0$ a procedure to estimate x_0 given by Kiefer and Wolfowitz [4] is as follows:

$$x_{n+1}=x_n+\frac{a_n}{c_n}(y_{2n}-y_{2n-1}) , \tag{2}$$

where $\{a_n\}$ and $\{c_n\}$ are sequences of positive numbers and where y_{2n} and y_{2n-1} are random variables which are conditionally independently distributed according to $F_{x_n+c_n}$ and $F_{x_n-c_n}$, respectively, given $x_1,\cdots,x_n$, $y_1,\cdots,y_{2n-2}$. The convergence of x_n to x_0 with probability one under some conditions is proved by Blum [1], Dvoretzky [3], Burkholder [2] and et al. As is shown in (1) and (2), the characteristic of $M(x)$ is made the best use of and then the above methods are thought not to be robust to the condition of M. It means that the stochastic approximation method is used seldom, if ever, in other cases.

From this point of view, a random observation process is proposed to estimate the location x_0. The estimation procedure is defined successively as follows, beginning with an initial value $x_1\in[0,1)$. For any positive integer n let $x_1, \cdots, x_n$ be the levels of observation chosen up to the nth stage and let $y_1, \cdots, y_n$ be the observed values at the respective levels. Using the x's and y's, we first construct an estimate M_n of the function M and then a probability distribution β_n over the interval $[0, 1)$. A subsequent level x_{n+1} is chosen by a chance mechanism with the distribution β_n and a corresponding value y_{n+1} is observed according to the distribution $F_{x_{n+1}}$. A reasonable criterion for the procedure may be that the distribution β_n converges to x_0 in law with probability one. The convergence is proved in Section 4 in case that the root of the equation $M(x)=a$ is to be estimated, while it will be discussed in the forthcoming paper [8] for the problem of estimating the maximum point of $M(x)$.

2. Notations and formulation

Let $\{F_x;\ 0 \leqq x < 1\}$ be a family of probability distributions on R^1 with finite means, say $M(x)$, such that for any Borel set B of R^1 $F_x(B)$ is measurable. For every non-negative integer m we define a half-open interval B_{mi} as

$$B_{mi} = \left[\frac{i-1}{2^m}, \frac{i}{2^m}\right) \qquad i=1, 2, \cdots, 2^m .$$

Let $\{k(m);\ m=0, 1, \cdots\}$ be a sequence of integers satisfying the following condition:

$$(3) \qquad k(m) \text{ is strictly increasing and } k(0)=1 .$$

Let Ω be the Cartesian product $\prod_{i=1}^{\infty} (\mathcal{X}_i, \mathcal{Y}_i)$, where $\mathcal{X}_i = [0, 1)$ and $\mathcal{Y}_i = R^1$. Let $\mathfrak{A}$ be its Borel σ-field. For any $x \in [0, 1)$, $\omega = (x_1, y_1; x_2, y_2; \cdots) \in \Omega$ and $n=1, 2, \cdots$ we define $X_n(\omega) = x_n$, $Y_n(\omega) = y_n$ and $N_n(x, \omega) = k(m_0)$, where

$$m_0 = \max \{m;\ \#[j;\ X_j(\omega) \in B_{mi},\ x \in B_{mi},\ j=1, \cdots, n] \geqq k(m)\} ,$$

the symbol $\#[A]$ denoting the number of the elements of the set A. The number m_0 is determined uniquely. We denote $B_n(x, \omega) = B_{m_0 i_0}$, where $N_n(x, \omega) = k(m_0)$ and $x \in B_{m_0 i_0}$ and let $C_n(x, \omega)$ be the set of $X_j(\omega)$ belonging to $B_n(x, \omega)$ with the smallest subscript, $k(m_0)$ in number. Define a random function on $[0, 1)$

$$(4) \qquad M_n(x, \omega) = \frac{1}{N_n(x, \omega)} \sum_{X_j(\omega) \in C_n(x, \omega)} Y_j(\omega) ,$$

which may serve as an approximation of $M(x)$ and put

(5) $$Z_n(\omega)=Y_n(\omega)-M(X_n(\omega))$$

and $$M_n'(x, \omega)=\frac{1}{N_n(x, \omega)}\sum_{X_j(\omega)\in C_n(x,\omega)} Z_j(\omega) .$$

It follows from the measurability of $F_x(B)$ for every Borel set B of R^1 that $M(x)$ and hence $Z_n(\omega)$ are measurable. Furthermore let

(6) $$\alpha_n(x, \omega)=h\{N_n(x, \omega)\}\exp\{-N_n(x, \omega)|M_n(x, \omega)-a|^b\} ,$$

where b is a positive constant and $h(m)>0$ for every $m=1, 2,\cdots$. Finally define a density function on $[0, 1)$ by

(7) $$\beta_n(x, \omega)=\frac{\alpha_n(x, \omega)}{\int_0^1 \alpha_n(t, \omega)dt}$$

and the corresponding probability distribution $P_{\beta_n(\omega)}$ on $[0, 1)$.

Since it can be easily proved from the measurability of $F_x(B)$ that $P_{\beta_n(\omega)}(A)$ is measurable for every Borel set A of $[0, 1)$, we can consider the probability measure P on $(\Omega, \mathfrak{A})$ having the following properties (i) and (ii) for every $n=1, 2,\cdots$.

(i) For every Borel set A of $[0, 1)$

(8) $$P(A'|X_i(\omega)=x_i,\ Y_i(\omega)=y_i,\ i=1,\cdots, n)=P_{\beta_n(\omega)}(A) ,$$

where $A'=X_{n+1}^{-1}(A)$.

(ii) For every Borel set B of R^1

(9) $$P(B'|X_i(\omega)=x_i,\ Y_j(\omega)=y_j,\ i=1,\cdots, n,\ j=1,\cdots, n-1)=F_{x_n}(B) ,$$

where $B'=Y_n^{-1}(B)$.

3. Lemmas

The arguments in this section are based on the probability space $(\Omega, \mathfrak{A}, P)$ introduced in Section 2.

LEMMA 1. *Under the following two conditions:*

(10) $$\sum_{m=0}^{\infty} 2^m k(m)^{2-b/b}<\infty \qquad \text{and} \qquad b>2 ,$$

(11) *the variances of random variables with the distribution F_x are uniformly bounded by $V<\infty$,*

we have

$$\overline{\lim_{n\to\infty}} \sup_{0\le x<1} |M_n'(x, \omega)|^b N_n(x, \omega) < \infty \text{ w.p. } 1 . \tag{12}$$

PROOF. By the definition of $M_n(x, \omega)$ we have

$$\begin{aligned}(13)\qquad P_\infty &\equiv P\left\{\omega;\ \overline{\lim_{n\to\infty}} \sup_{0\le x<1} |M_n'(x, \omega)|^b N_n(x, \omega) = \infty\right\} \\ &= \lim_{K\to\infty} P\{\omega;\ \text{There exist some } m,\ i,\ n \text{ and } x \text{ such that} \\ &\qquad B_{mi} = B_n(x, \omega) \text{ and that } |M_n'(x, \omega)| N_n(x, \omega)^{1/b} > K\} \\ &\le \lim_{K\to\infty} \sum_{m=0}^{\infty} \sum_{i=1}^{2^m} P_{mi} ,\end{aligned}$$

where

$$\begin{aligned}P_{mi} = P\{\omega;\ &\text{There exist some } x \text{ and } n \text{ such that} \\ &B_{mi} = B_n(x,\omega) \text{ and } |M_n'(x, \omega)| N_n(x, \omega)^{1/b} > K\} .\end{aligned}$$

Now we fix i and m and for every $\omega \in \Omega$ and for every positive integer l, let $t(l, \omega)$ be the lth smallest index j that $X_j(\omega) \in B_{mi}$. When such index does not exist we denote $t(l, \omega) = \infty$. We define

$$Z'(l, \omega) = \begin{cases} Z_{t(l,\omega)}(\omega) & \text{if } t(l, \omega) \ne \infty \\ 0 & \text{if } t(l, \omega) = \infty . \end{cases}$$

Then

$$\begin{aligned}(14)\qquad P_{mi} &= P\Big\{\omega;\ \#[j;\ X_j(\omega) \in B_{mi}] \ge k(m) \text{ and} \\ &\qquad\qquad \Big|\sum_{X_j(\omega) \in C_n(x,\omega)} Z_j(\omega)\Big| > Kk(m)^{1-1/b}\Big\} \\ &\le P\left\{\omega;\ \left|\sum_{l=1}^{k(m)} Z'(l, \omega)\right| > Kk(m)^{1-1/b}\right\} .\end{aligned}$$

Moreover by Markov's inequality we have

$$\begin{aligned}(15)\qquad P_{mi} &\le K^{-2} k(m)^{-2+2/b} E\left\{\sum_{l=1}^{k(m)} Z'(l, \omega)\right\}^2 \\ &= K^{-2} k(m)^{-2+2/b} \left\{\sum_{l=1}^{k(m)} EZ'^2(l, \omega) + 2 \sum_{l<l'} Z'(l, \omega) Z'(l', \omega)\right\} .\end{aligned}$$

But (9) and (11) imply

$$EZ'^2(l, \omega) = E\{E(Z'^2(l, \omega) | t(l, \omega) = t,\ X_t(\omega) = x)\} \le V$$

and for every $l < l'$

$$\begin{aligned}EZ'(l, \omega) Z'(l', \omega) = E\{E(Z'(l, \omega) Z'(l', \omega) | &t(l, \omega) = t,\ t(l', \omega) = t', \\ &X_i(\omega) = x_i,\ i = 1, \cdots, t',\ Y_j(\omega) = y_j, \\ &j = 1, \cdots, t'-1)\} = 0 .\end{aligned}$$

Then

$$P_{mi} \leqq K^{-2}k(m)^{-1+2/b}V . \tag{16}$$

Thus from (13) and (16) we have

$$P_{\infty} \leqq \lim_{K\to\infty} \sum_{m=0}^{\infty} \sum_{i=1}^{2^m} K^{-2}k(m)^{-1+2/b}V = \lim_{K\to\infty} K^{-2} \sum_{m=0}^{\infty} 2^m k(m)^{-1+2/b}V = 0 ,$$

where the last equality follows from (10). Hence Lemma 1.

LEMMA 2. *Suppose that the conditions* (10) *and* (11) *are satisfied and further that $h(m)$ and $M(x)$ satisfy the following conditions:*

(17) *for every $\alpha > 0$ the sequence $\{h(m)\exp(-\alpha m),\ m=1, 2, \cdots\}$ is bounded ,*

$$\begin{aligned} &(M(x)-a)(x-x_0)>0 \qquad \text{for any } x \neq x_0 , \qquad \text{and} \\ &\inf_{|x-x_0|>\varepsilon} |M(x)-a| \neq 0 \qquad \text{for any } \varepsilon>0 . \end{aligned} \tag{18}$$

Then for every $\varepsilon>0$ the set $\{\alpha_n(x, \omega);\ n=1, 2, \cdots, |x-x_0|>\varepsilon\}$ is bounded with probability one.

PROOF. We fix $\omega \in \Omega$ which satisfies (12) and let $C=C(\omega)$ satisfy

$$|M_n'(x, \omega)|^b N_n(x, \omega) < C \qquad \text{for any } n=1, 2, \cdots,\ 0 \leqq x < 1 .$$

From the definition of $B_n(x, \omega)$ it follows that for every $\varepsilon>0$ only finite many of the sets $B_n(x, \omega)$ with $n=1, 2, \cdots$ and $x<x_0-\varepsilon$ satisfy that $x_0-(\varepsilon/2) \in B_n(x, \omega)$. Then we can choose positive integers n_0 and p and the half open intervals $E_1, E_2, \cdots, E_p$ such that

(i) $\alpha_n(x, \omega)$ is constant with respect to $n>n_0$ and to $x \in E_j$ $j=1, 2, \cdots, p$ and

(ii) for every $x \in \left(\bigcup_{j=1}^{p} E_j\right)^c \cap [0, x_0-\varepsilon)$ and for every $n>n_0$ the set $C_n(x, \omega)$ does not contain any $X_i(\omega)>x_0-(\varepsilon/2)$.

Hence we have only to prove that $\alpha_n(x, \omega)$ is bounded with respect to $x \in \left(\bigcup_{j=1}^{p} E_j\right)^c \cap [0, x_0-\varepsilon)$ and to $n>n_0$. From the assumption (18) there exists some $\varepsilon'>0$ such that

$$M(x)-a < -\varepsilon' \qquad \text{for any } x<x_0-(\varepsilon/2) . \tag{19}$$

And we can choose an N_0 such that for any $N_n(x, \omega)>N_0$

$$|M_n'(x, \omega)| < C^{1/b}(N_n(x, \omega))^{-1/b} < \varepsilon'/2 . \tag{20}$$

From (19) and (20) it follows

$$M_n(x,\omega)-a=\frac{1}{N_n(x,\omega)}\left\{\sum_{X_j(\omega)\in C_n(x,\omega)} M(X_j(\omega))-a\right\}+M_n'(x,\omega)$$
$$\leqq -\varepsilon'/2 \qquad \text{for } N_n(x,\omega)>N_0\,.$$

On the other hand for $N_n(x,\omega)\leqq N_0$, $\alpha_n(x,\omega)$ takes only finite distinct values. From the above facts and (17), $\alpha_n(x,\omega)$ is bounded on $x<x_0-\varepsilon$ and $n=1,2,\cdots$. The boundedness of $\alpha_n(x,\omega)$ on $x>x_0+\varepsilon$ is proved in an entirely similar way and then this completes the proof.

Before stating the next lemma we need two notations. $B_n'(x_0,\omega)$ and $N_n'(x_0,\omega)$ are defined as $B_n(x_0-\varepsilon,\omega)$ and $N_n(x_0-\varepsilon,\omega)$ for sufficiently small $\varepsilon>0$. Since for every x' $B_n(x,\omega)$ and $N_n(x,\omega)$ are constant on some interval containing x', the above notations are well defined. Note that $B_n'(x_0,\omega)\not\equiv B_n(x_0,\omega)$ implies that x_0 can be written as $x_0=q/2^p$ for some integers p and q.

LEMMA 3. *Under the same assumptions as in Lemma 2, $N_n(x_0,\omega)$ or $N_n'(x_0,\omega)$ diverges to infinity as $n\to\infty$ with probability one.*

PROOF. Let E be an open neighbourhood of x_0 and let

$$B_i=\{\omega;\ X_i(\omega)\in E\} \qquad i=1,2,\cdots\,.$$

We denote $\mathfrak{B}_n$ the sub-σ-field of $\mathfrak{A}$ generated by $B_1,B_2,\cdots,B_n$. Then by extended Borel's zero-one law (e.g. Loève [5]) the two sets $\left\{\omega;\ \sum_{n=1}^{\infty} P^{\mathfrak{B}_{n-1}}B_n<\infty\right\}$ and $\left\{\omega;\ \sum_{n=1}^{\infty} I_{B_n}<\infty\right\}$ are equivalent, where I_{B_n} is an indicator function of B_n. Thus for every positive c

$$\omega\in\Omega_c=\left\{\omega;\ \varliminf_{n\to\infty}\int_E \beta_n(x,\omega)dx>\frac{1}{c}\right\}$$

implies $\sum_{n=1}^{\infty} I_{B_n}(\omega)=\infty$ with probability one. Otherwise from Lemma 2 we see that if $\omega\notin\bigcup_{c=1}^{\infty}\Omega_c$ there are infinite $X_i(\omega)\in E$ with probability one. Hence Lemma 3 follows.

LEMMA 4. *In addition to the assumptions of Lemma 2 we further impose the following conditions:*

(21) $\qquad 2^{-m}h(k(m))\to\infty \qquad$ *as* $m\to\infty$,

(22) $\qquad \{k(m)2^{-bm};\ m=1,2,\cdots\} \qquad$ *is bounded,*

(23) $\qquad \varlimsup_{x\to x_0}\left|\dfrac{M(x)-a}{x-x_0}\right|<A<\infty \qquad$ *for some positive A.*

Then

$$\int \alpha_n(x, \omega)dx \to \infty \qquad as\ n \to \infty\ w.p.\ 1\,. \tag{24}$$

PROOF. By Lemma 3 $N_n(x_0, \omega)$ or $N_n'(x_0, \omega)$ diverges to infinity with probability one, but we shall prove this lemma in only the case when $N_n(x_0, \omega) \to \infty$ as its proof is similar in other case. Now it holds that $\alpha_n(x, \omega) = \alpha_n(x_0, \omega)$ for any x belonging to one of the two sub-intervals of $B_n(x_0, \omega)$ of half size which contains x_0. Let

$$B_n(x_0, \omega) = B_{mi}\,.$$

Then

$$\begin{aligned}(25) \quad & \int_0^1 \alpha_n(x, \omega)dx \\ & \geqq \int_{B_{m_i}} \alpha_n(x, \omega)dx \\ & \geqq 2^{-1-m} h(k(m)) \exp\{-k(m)\,|\,M_n(x_0, \omega) - a\,|^b\} \\ & \geqq 2^{-1-m} h(k(m)) \exp\left\{-k(m) 2^{b-1} \left| \sum_{X_j(\omega) \in C_n(x_0, \omega)} \frac{M(X_j(\omega)) - a}{k(m)} \right|^b \right\} \\ & \quad \cdot \exp\left\{-k(m) 2^{b-1} \left| \sum_{X_j(\omega) \in C_n(x_0, \omega)} \frac{Z_j(\omega)}{k(m)} \right|^b \right\}\,.\end{aligned}$$

The first factor in the last member of this chain of inequalities diverges to infinity by (21) since $N_n(x_0, \omega) \to \infty$ implies $m \to \infty$. The second and the third are bounded away from zero by (22), (23) and Lemma 1. Then Lemma 4 follows.

Finally we show an example of $k(m)$ and $h(m)$ satisfying the conditions (3), (10), (17), (21) and (22) when $b > 3$. Let $k'(m) = (2^m m^2)^{b/b-2}$ and $h(m) = m^c$, where $c > 1 - 2/b$. Then $2^m k'(m)^{-1+2/b} = m^{-2}$ implies (10) with $k(m)$ replaced by $k'(m)$. It is easy to verify (17) and (22). And

$$2^{-m} h(k'(m)) = 2^{m(bc-b+2)/(b-2)} m^{2bc/(b-2)}$$

implies (21). Now let $k(0) = 1$ and let $k(m)$ be the integer part of $k'(m)$ for $m = 1, 2, \cdots$. Then it is easily seen that $k(m)$ and $h(m)$ satisfy all the required conditions.

4. Main theorems

THEOREM 1. *Let the probability space* $(\Omega, \mathfrak{A}, P)$ *have the properties* (8) *and* (9). *And let* $k(m)$ *and* $h(m)$ *satisfy the conditions* (3), (10), (17), (21) *and* (22). *If* F_x *and* $M(x)$ *satisfy* (11), (18) *and* (23), *then*

$$(26)\qquad \lim_{n\to\infty}\int_E \beta_n(x,\omega)dx=1 \qquad \textit{with probability one}$$

for every open interval E containing x_0.

PROOF. Let E be an open interval containing x_0. By definition of $\beta_n(x, \omega)$

$$(27)\qquad \int_{E^c}\beta_n(x,\omega)dx=\frac{\int_{E^c}\alpha_n(x,\omega)dx}{\int_0^1\alpha_n(x,\omega)dx}\,.$$

As $n\to\infty$ the numerator of the right hand-side is bounded with probability one by Lemma 2 and the denominator diverges to infinity with probability one by Lemma 4. The above two facts imply Theorem 1.

To generalize the theorem to the case when $M(x)=a$ has many roots we postulate the following conditions (28) and (29):

(28) $\inf\{|M(x)-a|;\ \min_{i=1,\cdots,p}|x-x_i|>\varepsilon\}>0$ for every $\varepsilon>0$ and the signature of $M(x)-a$ keeps the same on each interval (x_i, x_{i+1}) for every $i=0,1,\cdots,p$, where $x_0=0$ and $x_{p+1}=1$.

$$(29)\qquad \overline{\lim_{x\to x_i}}\left|\frac{M(x)-a}{x-x_i}\right|<A<\infty \qquad \text{for some } A\ \ i=1,\cdots,p\,.$$

THEOREM 2. *In the assumptions in Theorem* 1, *replace* (18) *and* (23) *by* (28) *and* (29). *Then*

$$\lim_{n\to\infty}\int_E \beta_n(x,\omega)dx=1 \qquad \textit{with probability one}$$

for every $E=\{x;\ \min_{i=1,\cdots,p}|x-x_i|<\varepsilon\}$ with $\varepsilon>0$.

PROOF. Under the new assumptions all that has to be done is to prove the following two lemmas:

LEMMA 2′. *For every $\varepsilon>0$ $\{\alpha_n(x,\omega);\ n=1,2,\cdots.\ |x-x_i|>\varepsilon,\ i=1,\cdots,p\}$ are bounded with probability one.*

LEMMA 3′. *Some of the set $\{N_n(x_i,\omega), N_n'(x_i,\omega)\}$ $i=1,\cdots,p$ diverge to infinity with probability one.*

They can be proved similarly as the corresponding lemmas stated before and details are omitted.

Remark 1. The theorems can be extended to the case when a statistician takes several levels at one time which are decided according to the distribution P_{β_n} and gets an observation for each level.

Remark 2. The expression (6) for the function $\alpha_n(x)$ is not essential. There may exist alternative forms for which Lemma 2 and Lemma 4 hold true.

Remark 3. The results in this paper are valid even if the parameter space of F_x is (0, 1), [0, 1] or (0, 1].

Acknowledgement

I wish to thank Professor M. Okamoto for his help and encouragement in the preparation of this paper.

NARA MEDICAL COLLEGE

REFERENCES

[1] Blum, J. R. (1954). Approximation methods which converge with probability one, *Ann. Math. Statist.*, **25**, 382-386.
[2] Burkholder, D. L. (1956). On a class of stochastic approximation processes, *Ann. Math. Statist.*, **27**, 1044-1059.
[3] Dvoretzky, A. (1956). On stochastic approximation, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, *University of California Press*, **1**, 39-55.
[4] Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.*, **23**, 462-466.
[5] Loeve, M. (1955). *Probability Theory*, Van Nostrand, New York.
[6] Robbins, H. and Monro, S. (1951). A stochastic approximation method, *Ann. Math. Statist.*, **22**, 400-407.
[7] Schmetterer, L. (1960). Stochastic approximation, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, *University of California Press*, **1**, 587-609.
[8] Yamazoe, S. On random observation processes for stochastic approximation. (To be submitted.)