

El Sayed Sorour

An extension of Driml-Nedoma continuous stochastic approximation procedure

*Kybernetika*, Vol. 14 (1978), No. 1, (18)--27

Persistent URL: <http://dml.cz/dmlcz/125402>

## Terms of use:

© Institute of Information Theory and Automation AS CR, 1978

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these

*Terms of use.*



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*  
<http://project.dml.cz>

# An Extension of Driml-Nedoma Continuous Stochastic Approximation Procedure

EL SAYED SOROUR

We extend the work of Driml-Nedoma to the case where the regression function has several roots and the noise also is dependent on  $x$ . We also show how this procedure can be used to evaluate the minimum of a regression function (Kiefer-Wolfowitz procedure).

## 1. INTRODUCTION

Driml and Nedoma [1], have proposed the following differential equation

$$(1.1) \quad \frac{dX(t)}{dt} = -\frac{1}{t}(r(X(t)) + \zeta(t))$$

as a continuous analogy of Robbins-Monro stochastic approximation procedure [5]. They have proved that under certain general conditions on  $r(x)$  and  $\zeta(t)$ , the solution of (1.1) converges with probability one to the unique root of the regression function  $r(x)$ .

In this paper, we shall be concerned with the extension of this procedure to the case where the regression function  $r(x)$  has several (finite number of) roots and we allow the noise  $\zeta(t)$  to be also dependent on  $x$ .

We also show that this procedure can be applied to estimate the minimum (maximum) of the regression function (Kiefer-Wolfowitz procedure) and the convergence with probability one of the procedure is proved.

## 2. BASIC ASSUMPTIONS AND NOTATIONS

All random variables are supposed to be defined on a complete probability space  $(\Omega, \mathfrak{F}, P)$ . Relations between random variables are meant with probability one.  $E$  denotes the expectation. The real line is denoted by  $R$ .  $\rho$  denotes the ordinary metric on  $R$  and  $\emptyset$  the empty set.

The following assumptions are supposed to be satisfied in the sequel:

- (i) the regression function  $r(x)$  is real-valued continuous;
- (ii)  $\zeta(t, x)$  is a real stochastic process, its sample functions are continuous with respect to  $t$  and  $x$  simultaneously with probability one;
- (iii) the functions  $a(t)$  and  $c(t)$  are defined as follows.

$$(2.1) \quad \begin{aligned} a(t) &= 1 \quad \text{for } 0 \leq t \leq 1, \\ &= \frac{1}{t} \quad \text{for } t > 1; \\ c(t) &= 1 \quad \text{for } 0 \leq t \leq 1, \\ &= \frac{1}{t^\gamma} \quad \text{for } t > 1, \quad \gamma > 0. \end{aligned}$$

### 3. CONDITIONS AND PRELIMINARIES

The following conditions will be needed as referred to.

*Conditions on the regression function  $r(x)$*

- A 1:**  $r(x)$  has several (finite number of) roots. Let the roots be ordered as  $-\infty < \theta_1 < \theta_2 < \dots < \theta_n < \infty$ , and let us have for every  $\delta > 0$

$$\sup_{-\infty < x < \theta_1 - \delta} r(x) < 0; \quad \inf_{\theta_n + \delta < x < \infty} r(x) > 0.$$

- A 2:**  $r(x)$  has several (finite number of) stationary points (roots of its derivative  $\bar{r}(x)$ ).  $\bar{r}(x)$  exists, and it is continuous and satisfies Lipschitz condition. Let the roots of  $\bar{r}(x)$  be ordered as  $-\infty < \theta_1 < \theta_2 \dots < \theta_n < \infty$ , and let us have for every  $\delta > 0$

$$\sup_{-\infty < x < \theta_1 - \delta} \bar{r}(x) < 0; \quad \inf_{\theta_n + \delta < x < \infty} \bar{r}(x) > 0.$$

*Conditions on the noise  $\zeta(t, x)$*

- B 1:**  $h_1(t) \leq \zeta(t, x) \leq h_2(t)$  for every  $t$  and  $x$  with probability one, where  $h_1(t)$  and  $h_2(t)$  are real continuous stochastic processes.

- B 2:**  $\lim_{t \rightarrow \infty} (1/t) \int_0^t h_i(\tau) d\tau = 0$  with probability one,  $i = 1, 2$ .

Let  $C_i(t, t')$ ,  $i = 1, 2$ , be the covariance functions of  $h_i(t)$ .

- C 1:**  $C_i(t, t)$  is a continuous function,  $i = 1, 2$ ; let  $c$ ,  $\alpha$  and  $k$  be finite positive constants, then

$$C_i(t, t) \leq c \quad \text{and} \quad \frac{1}{t^2} \int_0^t \int_0^t |C_i(\tau, \tau')| d\tau d\tau' \leq \frac{k}{t^\alpha};$$

$i = 1, 2$ , for large  $t$ .

20 **Remark 3.1.** The condition **C 1** is satisfied especially if  $h_i(t)$ ,  $i = 1, 2$ , are real stationary processes (wide sense) with the covariance functions

$$C_i(s, s + t) = R_i(t)$$

satisfying

$$\int_0^\infty |R_i(t)| dt < \infty; \quad i = 1, 2;$$

in this case **C 1** is satisfied with  $\alpha = 1$ .

The lemma, due to M. Driml and J. Nedoma is used repeatedly. The second part of the lemma is as follows.

**Lemma 3.1.** Let  $g(t)$  be a continuous real-valued function such that

$$(3.1) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t g(\tau) d\tau = m > 0$$

exists. Let  $\eta$  be a given number such that  $0 < \eta < m$  and let  $T > 1$  be so that for  $t \geq T$ , the inequality

$$(3.2) \quad \left| \frac{1}{t} \int_0^t g(\tau) d\tau - m \right| < \eta$$

holds. If  $v \geq u \geq T$ , then

$$(3.3) \quad \int_u^v a(\tau) g(\tau) d\tau > -2\eta + (m - \eta) \ln \frac{v}{u} \geq -2\eta.$$

#### 4. ALMOST SURE CONVERGENCE OF ROBBINS-MONRO PROCEDURE

Let  $X(t)$  be the solution of the differential equation

$$(4.1) \quad \frac{dX(t)}{dt} = -a(t) [r(X(t)) + \zeta(t, X(t))].$$

From the assumptions (i)–(iii), it follows that  $X(t)$  exists and that it is continuous with probability one.

**Theorem 4.1.** If **A 1**, **B 1** and **B 2** hold, then  $g(X(t), \Theta) \rightarrow 0$  with probability one for  $t \rightarrow \infty$ ,

where

$$\Theta = \{x : r(x) = 0\}.$$

Proof. Let us denote by  $\Omega_1$  the set  $\{\omega : \zeta(t, x) \text{ is continuous with respect to } t \text{ and } x \text{ simultaneously; } h_i(t), i = 1, 2, \text{ are continuous; for every } t \text{ and every } x h_1(t) \leq \zeta(t, x) \leq h_2(t); \lim_{t \rightarrow \infty} (1/t) \int_0^t h_i(\tau) d\tau = 0; i = 1, 2\}$ .

From the hypothesis of the theorem we conclude that  $P(\Omega_1) = 1$ .

First we shall prove that

$$\liminf_{t \rightarrow \infty} \varrho(X(t), \Theta) = 0$$

with probability one.

Let us suppose that it is not so. Let

$$(4.2) \quad \mathfrak{D} = \{\omega \in \Omega_1 : \liminf_{t \rightarrow \infty} \varrho(X(t), \Theta) > 0\};$$

then  $\mathfrak{D} \neq \emptyset$ . Taking  $\omega \in \mathfrak{D}$ , there are two numbers  $\eta > 0$  and  $t_1 \geq 1$  such that

$$(4.3) \quad \varrho(X(t), \Theta) > \eta \quad \text{for all } t \geq t_1.$$

Denote  $\theta_0 = -\infty$  and  $\theta_{n+1} = \infty$ .

From the continuity of  $X(t)$  and (4.3) there exists an index  $i_0 = 1, 2, \dots, n+1$  such that  $\theta_{i_0} - \theta_{i_0-1} > 2\eta$  and at the same time

$$(4.4) \quad X(t) \in (\theta_{i_0-1} + \eta, \theta_{i_0} - \eta) \quad \text{for all } t \geq t_1.$$

Thus we have two possibilities

$$(a) \quad X(t) \in (\theta_{i_0-1} + \eta, \theta_{i_0} - \eta), \quad r(X(t)) > 0, \quad t \geq t_1,$$

$$(b) \quad X(t) \in (\theta_{i_0-1} + \eta, \theta_{i_0} - \eta), \quad r(X(t)) < 0, \quad t \geq t_1.$$

We shall prove that a contradiction can be deduced from the case (a) (the same holds for the case (b), however, the proof being analogous will be omitted).

Defining

$$(4.5) \quad m = \inf_{x \in (\theta_{i_0-1} + \eta, \theta_{i_0} - \eta)} r(x),$$

then it is evident that  $m > 0$ .

Let us choose a number  $\Delta$  so that  $0 < \Delta < m$  and let  $t_2 > t_1 > 1$  be such a point that for all  $t \geq t_2$

$$(4.6) \quad \left| \frac{1}{t} \int_0^t h_1(\tau) d\tau \right| < \Delta.$$

Integrating both sides of the differential equation (4.1) from  $t_2$  to  $t$  and using (4.6) and (3.3) we conclude that

$$(4.7) \quad X(t) - X(t_2) = - \int_{t_2}^t a(\tau) [r(X(\tau)) + \zeta(\tau, X(\tau))] d\tau \leq$$

$$\leq - \int_{t_2}^t a(\tau) (m + h_1(\tau)) d\tau < 2\Delta - (m - \Delta) \ln \frac{t}{t_2}.$$

Thus we have  $\lim_{t \rightarrow \infty} X(t) = -\infty$  what contradicts (a), therefore  $\mathcal{D} = \emptyset$ . Thus

$$(4.8) \quad \liminf_{t \rightarrow \infty} \varrho(X(t), \Theta) = 0 \quad \text{with probability one.}$$

Next we shall show that

$$\limsup_{t \rightarrow \infty} \varrho(X(t), \Theta) = 0 \quad \text{with probability one.}$$

Let

$$(4.9) \quad H_i = \{\omega \in \Omega_1 : \liminf_{t \rightarrow \infty} \varrho(X(t), \theta_i) = 0\}, \quad i = 1, 2, \dots, n;$$

from (4.8)  $\mathbb{P}(\bigcup_{i=1}^n H_i) = 1$ .

Let  $k$  be an arbitrary number where  $k = 1, 2, \dots, n$  and let us denote

$$\mathcal{D}_k = \{\omega \in H_k : \limsup_{t \rightarrow \infty} \varrho(X(t), \theta_k) > 0\}.$$

Let  $\omega \in \mathcal{D}_k$ , then there exists a number  $\eta > 0$ , such that for every  $t$  there exists a point  $t_1 > t$  such that

$$(4.10) \quad \varrho(X(t_1), \theta_k) > \eta.$$

Denoting by  $L_k$  the length of the interval  $(\theta_{k-1}, \theta_k)$  and

$$(4.11) \quad \delta = \min\left(\frac{L_k}{2}, \frac{L_{k+1}}{2}, \eta\right),$$

from the continuity of  $X(t)$ , from (4.9) and (4.10) one of these assertions follows:

- (1) for each  $t$  there exists a point  $t_2 > t_1$  such that  $X(t_2) = \theta_k + \delta$ ;
- (2) for each  $t$  there exists a point  $t_2 > t_1$  such that  $X(t_2) = \theta_k - \delta$ .

Let us consider the assertion (1). First we assume that

$$(A) \quad (\theta_k, \theta_{k+1}) \subset \{x : r(x) > 0\}.$$

Defining

$$m = \min_{\delta/2 \leq x - \theta_k \leq \delta} r(x),$$

$$\Delta = \min(m, \delta/4),$$

there exists  $t_3 \geq t_2 > 1$  such that

$$(4.12) \quad \left| \frac{1}{t} \int_0^t h_1(\tau) d\tau \right| < \Delta \quad \text{for all } t \geq t_3.$$

From (4.9) there exists  $t_4 > t_3$  such that

$$(4.13) \quad X(t_4) = \theta_k + \frac{\delta}{2}.$$

From assertion (1) there exists  $t_6 > t_4$  such that

$$(4.14) \quad X(t_6) = \theta_k + \delta.$$

Let  $t_5 < t_6$  be the maximum time for which

$$(4.15) \quad X(t_5) = \theta_k + \frac{\delta}{2},$$

$t_5$  exists because of the continuity of  $X(t)$ , (4.9) and the assertion (1). Then

$$(4.16) \quad X(t) > \theta_k + \frac{\delta}{2} \quad \text{for } t_5 < t \leq t_6.$$

Integrating both sides of the differential equation (4.1) from  $t_5$  to  $t_6$ , then from (4.16) and (3.3), we have

$$(4.17) \quad \begin{aligned} \frac{\delta}{2} = X(t_6) - X(t_5) &= - \int_{t_5}^{t_6} a(\tau)(r(\tau) + \zeta(\tau, X(\tau))) d\tau \leq \\ &\leq - \int_{t_5}^{t_6} a(\tau)(m + h_1(\tau)) d\tau < 2\Delta \leq \frac{\delta}{2}, \end{aligned}$$

which is the desired contradiction. Thus  $\vartheta_k = \emptyset$ ,  $k = 1, 2, \dots, n$ , if the assertion (1) and the assumption (A) hold.

Analogously, the case

$$(B) \quad (\theta_k, \theta_{k+1}) \subset \{x : r(x) < 0\}$$

under the validity of assertion (1) can be treated.

The assertion (2) can be ruled out as (1). Thus  $\vartheta_k = \emptyset$  for an arbitrary index  $k = 1, 2, \dots, n$ , from which it follows that

$$\limsup_{t \rightarrow \infty} \varrho(X(t), \Theta) = 0 \quad \text{with probability one.}$$

This completes the proof of the theorem.

5. ALMOST SURE CONVERGENCE OF KIEFER AND WOLFOWITZ PROCEDURE

Let  $X(t)$  be the solution of the differential equation

$$(5.1) \quad \frac{dX(t)}{dt} = \frac{-a(t)}{2c(t)} [r(X(t) + c(t)) - r(X(t) - c(t)) + \zeta(t, X(t))].$$

From (i)–(iii),  $X(t)$  exists and is continuous with probability one.

To establish the almost sure convergence of the procedure we shall prove the following lemma. Its proof can be carried out in steps as in proving the Almost Sure Stability Theorem in [3], Section 34.7.

**Lemma 5.1.** Let  $h(t)$  be a real stochastic process with the covariance function  $C(t, t')$  satisfying C 1. Then for all  $0 \leq \gamma < \alpha/4$  we have

$$(5.2) \quad \frac{1}{t} \int_0^t u^\gamma h(u) du \rightarrow 0$$

with probability one for  $t \rightarrow \infty$ .

**Remark 5.1.** Let  $\zeta(t)$  be a stationary process (wide sense). Let  $R(t) = C(s + t, s)$ , then C 1 can be satisfied with  $\alpha = 1$  if

$$\int_0^\infty |R(t)| dt < \infty;$$

then  $\zeta(t)$  can be stationary Markov process (wide sense), because in this case

$$R(t) = e^{-\alpha t} \cos 2\pi\beta t R(0), \quad \alpha > 0, \quad t > 0,$$

$$R(0) < c.$$

**Proof of Lemma 5.1.** Let us define

$$Y(t) = \frac{1}{t} \int_0^t u^\gamma h(u) du.$$

We reduce the random function  $Y(t)$  to a sequence of random variables.

We consider a sequence  $m^a$ ,  $m = 1, 2, \dots$ , and  $a > 0$ . For

$$(5.3) \quad m^a \leq t < (m+1)^a,$$

we can set

$$(5.4) \quad \frac{t}{m^a} Y(t) = Y(m^a) + U(m^a, t), \quad U(m^a, t) = \frac{1}{m^a} \int_{m^a}^t u^\gamma h(u) du.$$



Since

25

$$Z(m^a) = \sup_{m^a \leq t < (m+1)^a} |U(m^a, t)| \leq \frac{1}{m^a} \int_{m^a}^{(m+1)^a} u^\gamma |h(u)| \, du,$$

we have

$$\begin{aligned} \mathbb{E}|Z(m^a)|^2 &\leq \frac{1}{m^{2a}} \int_{m^a}^{(m+1)^a} \int_{m^a}^{(m+1)^a} u^\gamma u'^\gamma \mathbb{E}|h(u)h(u')| \, du \, du' \leq \\ &\leq \left( \frac{1}{m^a} \int_{m^a}^{(m+1)^a} u^\gamma \sqrt{C(u, u)} \, du \right)^2. \end{aligned}$$

From **C 1** we can set

$$\sum_{m=1}^{\infty} \mathbb{E}|Z(m^a)|^2 \leq \sum_{m=1}^{\infty} \frac{c'^2}{m^{2a}} |(m+1)^{a(\gamma+1)} - m^{a(\gamma+1)}|^2 = \sum_{m=1}^{\infty} \frac{c'^2}{m^{2(1-a\gamma)}} \left(1 + O\left(\frac{1}{m}\right)\right),$$

$c'$  is constant and the series converges when

$$(5.5) \quad a\gamma < \frac{1}{2}.$$

Using Tchebichev's inequality, the Borel-Cantelli lemma and (5.3) we state

$$Z(m^a) \rightarrow 0 \quad \text{with probability one for } t \rightarrow \infty.$$

Thus

$$(5.6) \quad Y(t) - \frac{m^a}{t} Y(m^a) \rightarrow 0 \quad \text{with probability one for } t \rightarrow \infty.$$

Next, we put  $p = m^a$ ,  $m = 1, 2, \dots$ ; then

$$Y(p) = \frac{1}{p} \int_0^p u^\gamma h(u) \, du.$$

The second condition in **C 1** yields

$$\begin{aligned} \sum_p \mathbb{E}|Y(p)|^2 &= \sum_p \frac{1}{p^2} \int_0^p \int_0^p u^\gamma C(u, u') u'^\gamma \, du \, du' \leq \\ &\leq \sum_p \frac{1}{p^2} \int_0^p \int_0^p |C(u, u')| u^\gamma u'^\gamma \, du \, du' \leq \sum_p \frac{K}{p^{2-2\gamma}} = \sum_m \frac{1}{m^{a(\alpha-2\gamma)}} \end{aligned}$$

which is convergent when

$$(5.7) \quad a(\alpha - 2\gamma) > 1.$$

To complete the proof, it is sufficient to show that there exists an  $a > 0$  which satisfies (5.5) and (5.7). Put  $\varrho_1 = \alpha - 4\gamma$ . From the hypothesis of the lemma  $\varrho_1 > 0$ .

We can write (5.7) as  $a > 1/(2\gamma + \varrho_1)$ , from this and (5.5), we can put

$$\frac{1}{2\gamma + \varrho_1} < a < \frac{1}{2\gamma}$$

which completes the proof of the lemma.

**Theorem 5.1.** If **A 2**, **B 1** and **C 1** hold, then for all  $0 \leq \gamma < \alpha/4$  we have

$$\varrho(X(t), \Theta) \rightarrow 0$$

with probability one for  $t \rightarrow \infty$ , where

$$\Theta = \{x : \bar{r}(x) = 0\}.$$

*Proof.* The differential equation (5.1) can be written as

$$(5.9) \quad \frac{dX(t)}{dt} = -a(t)(\bar{r}(X(t)) + \zeta^*(t, X(t)))$$

where

$$\zeta^* = \zeta_1 + \zeta_2,$$

$$(5.10) \quad \zeta_1 = \frac{1}{2c(t)} (r(X(t) + c(t)) - r(X(t) - c(t)) - 2c(t) \bar{r}(X(t)))$$

$$(5.11) \quad \zeta_2 = \frac{1}{2c(t)} \zeta(t, X(t)).$$

It is clear that (5.9) is the same as (4.1) when  $r(x)$  and  $\zeta(t, x)$  in (4.1) are replaced by  $\bar{r}(x)$  and  $\zeta^*(t, x)$ . Therefore for the proof it is sufficient to show that the conditions of Theorem 4.1 are satisfied.

From **A 2** it is evident that **A 1** is satisfied with respect to  $\bar{r}(x)$ . The only conditions which are not directly stated here, are **B<sub>2</sub>** for  $\zeta_2$  and **B<sub>1</sub>**, **B<sub>2</sub>** for  $\zeta_1$ .

From (5.10) we have

$$\zeta_1 = \bar{r}(X(t) + \delta c(t)) - \bar{r}(X(t)), \delta \in [-1, 1],$$

then by **A 2**

$$|\zeta_1| \leq 2c(t);$$

then from (1.2), **B 1** and **B 2** are satisfied. From Lemma 5.1 and **C 1**  $\zeta_2$  satisfies **B 2**, which completes the proof of the theorem.

## ACKNOWLEDGEMENT

The author is grateful to Dr. M. Driml and Dr. V. Dupač for helpful discussions during the preparation of this work.

(Received July 26, 1977.)

## REFERENCES

---

- [1] M. Driml, J. Nedoma: Stochastic approximations for continuous random process. Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. Czechoslovak Academy of Sciences, Prague 1960, 145—158.
- [2] J. Kiefer, J. Wolfowitz: Stochastic estimation of the maximum of a regression function. The Annals of Mathematical Statistics 23 (1952), 462—466.
- [3] M. Loève: Probability theory. Van Nostrand, New York 1960.
- [4] M. B. Nevelson and R. Z. Hasminskij: Stochastic Approximation and Recurrent Estimations Nauka, Moscow, 1972. (In Russian.)
- [5] H. Robbins, S. Monro: A stochastic approximation method. The Annals of Mathematical statistics 22 (1951), 400—407.
- [6] D. J. Sackrison: A continuous Kiefer-Wolfowitz procedure for random processes. The Annals of Mathematical Statistics 35 (1964), 590—599.

*RNDr. Dipl. Eng. El Sayed El Sayed Farag Sorour, CSc., 8 - Ebn Marawan - Hammamat El Koppa, Cairo. Egypt.*