A GENERALIZATION OF DVORETZKY'S STOCHASTIC APPROXIMATION THEOREM

Neculai CRÎŞMARU

"G. Bacovia" University, Bacau, Romania (crismaru_nicolae@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.13

Abstract: *This presentation contains a generalization of one of Dvoretzk's well-known stochastic approximation theorem. This generalization is founded on the extension of one of Derman and Sacks's lemmas wich was the basis of a result given by the two authors in relation to Dvoretzky's stochastic approximation theorem.*

Keywords: *stochastic approximation theorems, Dvoretzky' theorem, Derman-Sacks lema.*

1. INTRODUCTION

For some time, the stochastic approximation has become a subject of interest for many researchers in different fundamental and applied fields. There are two famous algorithms wich are very often utilized in the research of stochastic approximations (among other algorithms): the Robbins-Monro algorithm ([5], [6]) and the Kiefer-Wolfowitz algorithm ([4]). A short time after these stochastic approximation algorithms appeared, A. Dvoretzky ([3]) discovered an approximation algorithm wich generalizesthe two previously mentioned algorithms. As a result of the particular interest generated by this new algorithm, many researchers have started studying it. Two of these researchers are worth mentioning: Derman and Sacks ([2]), who found a demonstration (based on a technical lemma) wich is shorter and easier than that Dvoretzky's initial theorem, wich has a rather lengthy and more difficult to apprehend demonstration. This presentation contains a generalization of Dvoretzky's initial theorem realized by means of a generalization (among many others) of Derman and Sacks's lemma. Firstly, we will provide Dvoretzky's initial theorem without its demonstration and Derman and Sacks's lemma for a better understanding of the starting point of this research.

Dvoretzky's Theorem ([3], pp. 39-55)

Consider a probability space (Ω, K, P).

Let be $(\alpha_n)_{n\in N^*}, (\beta_n)_{n\in N^*}, (\gamma_n)_{n\in N^*}$ such that $\alpha_n > 0$, $\beta_n \geq 0$, $\gamma_n \geq 0$,

$\forall n \in N^*$, $\alpha_n \xrightarrow[n\to+\infty]{} 0$, $\sum_{n=1}^{+\infty} \beta_n < +\infty$, $\sum_{n=1}^{+\infty} \gamma_n = +\infty$. Let be θ a real number and

$(T_n)_{n\in N^*}$ a sequence of real and measurable functions wich satisfies:

$|T_n(r_1, r_2, \ldots, r_n) - \theta| \leq \max\{\alpha_n; (1+\beta_n)|r_n - \theta| - \gamma_n\}$, $\forall n \in (N^* - \{1\})$ for every real numbers $r_1, r_2, \ldots, r_n$.

Let be the sequences of random variables: $(X_n)_{n\in N^*}$, $(Y_n)_{n\in N^*}$, such that

$X_{n+1}(\omega) = T_n(X_1(\omega), X_2(\omega) \ldots, X_n(\omega)) + Y_n(\omega)$ (a.s.), $\forall \omega \in \Omega$,

$\forall n \in (N^* - \{1\})$.

Suppose that the following conditions hold:

$E[Y_1^2]<+\infty$, $\sum_{n=1}^{+\infty} E[Y_n^2]<+\infty$, and $E[Y_n \mid X_1,X_2,\ldots,X_n]=0$ (a.s.), $\forall n\in N^*$. In the conditons above there exist:

$E[(X_n-\theta)^2]\xrightarrow[n\to+\infty]{}0$ with the probability one, and

$X_n\xrightarrow[n\to+\infty]{}\theta$ with the probabilty one.

Lema (Derman şi Sacks)([2])

Let be $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$, $(x_n)_{n\in N^*}$,sequences of real numbers wich satisfy the following conditions:

$(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(x_n)_{n\in N^*}\subset[0,+\infty)$, $\lim_{n\to+\infty}(a_n)=0$; $\sum_{n=1}^{+\infty} b_n<+\infty$;

$\sum_{n=1}^{+\infty} c_n=+\infty$; $\sum_{n=1}^{+\infty}\delta_n<+\infty$ şi $\exists\ N_0\in N^*$, such that for every $n>N_0$ we have:

$x_{n+1}\le\max\{a_n;(1+b_n)x_n+\delta_n-c_n\}$. Hence $\lim_{n\to+\infty}(x_n)=0$.

2. THE MAIN RESULT

Before we provide the generalization of Dvoretzky's theorem, we wil offer a series of helping lemmas, some of wich are generalization of Derman and Sack's lemma.

Lemma 1.

Let be the sequences of real number: $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$ $(\delta_n)_{n\in N^*}$, $(x_n)_{n\in N^*}$, such that we have:

$0\le x_{n+2}\le\max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n x_n+d_n\},\forall n\in N^*$, (1)

$a_n, b_n, c_n, e_n, x_n\ge 0, \delta_n, a_n\le 1$, $\forall n\in N^*$, (2)

$a_n\to 0$, $\sum_{n=1}^{+\infty} b_n<+\infty$, $\sum_{n=1}^{+\infty} e_n<+\infty$, $\sum_{n=1}^{+\infty} c_n=+\infty$, $\sum_{n=1}^{+\infty}\delta_n<+\infty$, and

$d_n=\delta_n-c_n$, $\forall n\in N^*$, $d_n\in R$, (3)

$|\delta_n|\le c_n$, $\forall n\in N^*$, $x_1,x_2\in R$, such that ($x_2\ge 1$ or $x_1\ge 1$), (4)

note that: $K=\max\{x_1,x_2\}$, (5)

suppose that the following conditions holds: $b_n+e_{n-1}\le\frac{-d_{n-1}}{K}$, (6)

$\forall n\in(N^*-\{1\})$.

Hence $(x_n)_{n\in N^*}$ isbounded and $0\le x_n\le K$.

<u>PROOF</u>: From (3) and (4) it result that $d_n=\delta_n-c_n\le|\delta_n|-c_n\le 0$, so we have:

$d_n\le 0$, $\forall n\in N^*$ (7)

From (5) it result $K\ge 1$ (8)

For n=1 from (1) we obtain:

$0\le x_3\le\max\{a_2;(1+b_2)x_2+e_1x_1+d_1\}$, (9)

We have the inequality :

$(1+b_2)x_2+e_1x_1+d_1\le(1+b_2)K+e_1K+d_1\le K(1+b_2+e_1+\frac{d_1}{K})$ (10)

For n=2 from (6) we obtain: $b_2+e_1+\frac{d_1}{K}\le 0$ and from (10) we have:

$(1+b_2)x_2+e_1x_1+d_1\le K.$ (11)

From (2), (8), (9) and (11) it result: $0\le x_3\le \max\{a_2;K\}=K$, so we have:

$0\le x_3\le K$ (12)

By induction we have:

$(1+b_{n+1})x_{n+1}+e_n x_n+d_n\le K(1+b_{n+1}+e_n+\frac{d_n}{K}), \forall n\in N^*,$ (13)

in (6) we replace n with (n+1) and we obtain:

$b_{n+1}+e_n+\frac{d_n}{K}\le 0$ and from (13) it result: $(1+b_{n+1})x_{n+1}+e_n x_n+d_n\le K$, and from (1) and (2) we obtain:

$0\le x_{n+2}\le \max\{a_{n+1};K\}=K$, so $0\le x_n\le K$, $\forall n\in N^*$, so $(x_n)_{n\in N^*}$ is bounded and $0\le x_n\le K$ (q.e.d.).

Lemma 2.

Let be the sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ and $(x_n)_{n\in N^*}$, such that we have:

$0\le x_{n+2}\le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n x_n+d_n\}, \forall n\in N^*,$ (1)

$a_n, b_n, c_n, e_n, x_n\ge 0,\ \delta_n, a_n\le 1,\ \forall n\in N^*,$ (2)

$a_n\to 0, \sum_{n=1}^{+\infty} b_n<+\infty, \sum_{n=1}^{+\infty} e_n<+\infty, \sum_{n=1}^{+\infty} c_n=+\infty, \sum_{n=1}^{+\infty} \delta_n<+\infty,$ (3)

$d_n=\delta_n-c_n, \forall n\in N^*, d_n\in R,$ (4)

$|\delta_n|\le c_n, \forall n\in N^*,$ (5)

$x_1, x_2\in R$, such that ($x_2\ge 1$ or $x_1\ge 1$), and we note that:

$K=\max\{x_1,x_2\},$ (6)

Suppose that the following holds: $b_n+e_{n-1}\le \frac{-d_{n-1}}{K}, \forall n\in (N^*-\{1\}).$ (7)

Hence $x_n \xrightarrow[n\to+\infty]{} 0.$

<u>PROOF</u>: From the previous lemma, (lemma 1.) we have that the sequence $(x_n)_{n\in N^*}$ is bounded and

$0\le x_n\le K$ and from (1) we have:

$0\le x_{n+2}\le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n K+d_n\}, \forall n\in N^*,$ (8)

Note that:

$e_n K+d_n=\tilde{d}_n$ (9)

so we have $0\le x_{n+2}\le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+\tilde{d}_n\}, \forall n\in N^*$, and from (4) we have:

with $\tilde{d}_n=e_n K+\delta_n-c_n,\ \forall n\in N^*,$ (10)

Facem acum altă notaţie:

$e_n K+\delta_n=\tilde{\delta}_n$ (11)

Using these notes (10) becomes:

$0 \le x_{n+2} \le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+\tilde{\delta}_n - c_n\}, \forall\, n \in N^*$, with $\tilde{d}_n = \tilde{\delta}_n - c_n$, (12)

$\forall\, n \in N^*$

and from (3) it result:

$$\sum_{n=1}^{+\infty} \tilde{\delta}_n = \sum_{n=1}^{+\infty} (e_n K + \delta_n) = K\sum_{n=1}^{+\infty} e_n - \sum_{n=1}^{+\infty} \delta_n < +\infty, \quad (13)$$

At this stage all the hypotheses of Derman and Sacks's lemma ([2], pp. 602) or of lemma 1. in this paper, so $x_n \to 0$ (q.e.d.).

Lemma 3.

Let be the real sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ $(x_n)_{n\in N^*}$, such that we have:

$0 \le x_{n+2} \le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n x_n + d_n\}, \forall\, n \in N^*$, (1)

$a_n, b_n, c_n, e_n, x_n \ge 0$, $\delta_n, a_n \le 1$, $\forall\, n \in N^*$, (2)

$a_n \to 0$, $\sum_{n=1}^{+\infty} b_n < +\infty$, $\sum_{n=1}^{+\infty} e_n < +\infty$, $\sum_{n=1}^{+\infty} c_n = +\infty$, $\sum_{n=1}^{+\infty} \delta_n < +\infty$, (3)

$d_n = \delta_n - c_n$, $\forall\, n \in N^*$, $d_n \in R$, (4)

$(x_n)_{n\in N^*}$ is bounded, $0 \le x_n \le K$, $\forall\, n \in N^*$. (5)

Hence $x_n \xrightarrow[n\to+\infty]{} 0$.

<u>PROOF</u>: From (1) and (5) we have: $0 \le x_{n+2} \le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n K + d_n\}$, $\forall\, n \in N^*$,

Note that: $e_n K + d_n = \tilde{d}_n$ (6)

so we have $0 \le x_{n+2} \le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+\tilde{d}_n\}, \forall\, n \in N^*$,

with $\tilde{d}_n = e_n K + \delta_n - c_n$, $\forall\, n \in N^*$, (7)

we add another note: $e_n K + \delta_n = \tilde{\delta}_n$ (8)

Using these notes (7) becomes:

$0 \le x_{n+2} \le \max\{a_{n+1};(1+b_{n+1})x_{n+1}+\tilde{\delta}_n - c_n\}, \forall\, n \in N^*$, (9)

with $\tilde{d}_n = \tilde{\delta}_n - c_n$, $\forall\, n \in N^*$, in wich, from (3) results that:

$$\sum_{n=1}^{+\infty} \tilde{\delta}_n = \sum_{n=1}^{+\infty} (e_n K + \delta_n) = K\sum_{n=1}^{+\infty} e_n - \sum_{n=1}^{+\infty} \delta_n < +\infty.$$

At this stage all hypotheses of Derman and Sack's lemma hold ([2], pag. 602), so $x_n \to 0$ (q.e.d.).

Lemma 4.

Let be the real sequences $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$, $(d_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ and $(x_n)_{n\in N^*}$, such that we have (for a chosen and fixed K):

1) $(x_n)_{n\in N^*}$ is bounded, $0 \le x_n \le K$, $\forall\, n \in N^*$, $K>0$,

2) $0 \le x_{n+2} \le \min\{\max\{a_{n+1};(1+b_{n+1})x_{n+1}+e_n x_n + d_n\};K\}$, $\forall\, n \in N^*$,

3) $a_n, b_n, c_n, e_n, x_n \ge 0$, $\delta_n, a_n \le 1$, $\forall\, n \in N^*$,

4) $a_n \to 0$, $\sum_{n=1}^{+\infty} b_n <+\infty$, $\sum_{n=1}^{+\infty} e_n <+\infty$, $\sum_{n=1}^{+\infty} c_n =+\infty$, $\sum_{n=1}^{+\infty} \delta_n <+\infty$,

5) $d_n = \delta_n - c_n$, $\forall n \in N^*$, $d_n \in R$.

Hence $x_n \xrightarrow[n\to+\infty]{} 0$.

PROOF: See lemma 3.

We now provide the main result of this presentation, result wich is a generalization of one of Dvoretzky's stochastic approximation theorem.

Theorem

Let be sequences $(X_n)_{n\in N^*}$,$(T_n)_{n\in N^*}$,$(Y_n)_{n\in N^*}$, such that the following conditions hold : X_n, Y_n are random variables on a probability space (Ω, K, P),

and X_i are random variables, $i\in\{1,2\}$, and $T_n : R^n \to R$, X_1,X_2 are measurables applications such that (X_1 or X_2)≥ 1 (a.s.), and the following conditions hold:

$$X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n \text{ (a.s.)} \tag{1}$$

$$E[Y_n \mid X_1, X_2, \ldots, X_n] = 0 \text{ (a.s.)},\ \forall n \in N^*, \tag{2}$$

$$\sum_{n=1}^{+\infty} E[Y_n^2] <+\infty, \tag{3}$$

$$|T_n(x_1, x_2, \ldots, x_n)| \leq \max\{\alpha_n ; (1+\beta_n)|x_n| + e_{n-1}|x_{n-1}| - \gamma_n\}, \tag{4}$$

$\forall n \in (N^*-\{1\})$.

Let be the sequences $(\alpha_n)_{n\in N^*}$,$(\beta_n)_{n\in N^*}$,$(\gamma_n)_{n\in N^*}$, $(e_n)_{n\in N^*}$, such that:

$\alpha_n \xrightarrow[n\to+\infty]{} 0$, $\sum_{n=1}^{+\infty} \beta_n <+\infty$, $\sum_{n=1}^{+\infty} e_n <+\infty$, $\sum_{n=1}^{+\infty} \gamma_n =+\infty$, and

$$\alpha_n > 0,\ \beta_n \geq 0,\ \gamma_n \geq 0,\ e_n \geq 0,\ \forall n \in N^*,\ \alpha_n \leq 1, \tag{5}$$

$$|Y_n| \leq \gamma_n \text{ (a.s.). Note that } K = \max\{X_1, X_2\}, \tag{6}$$

and suppose that:

$$\beta_n + e_n \leq \frac{1}{K}(|Y_{n-1}| - \gamma_n) \text{ (a.s.)},\ \forall n \in (N^*-\{1\}. \tag{7}$$

So we have:

a)$(X_n)_{n\in N^*}$ is bounded (a.s.), $|X_n| \leq K$, (a.s.)$\forall n \in (N^*-\{1\}$ and

b)$X_n \xrightarrow[n\to+\infty]{} 0$ (a.s.).

PROOF:

As in the demonstration of theorem 1. from ([2], pp. 603), we suppose that:

$$\sum_{n=1}^{+\infty} \frac{1}{\alpha_n^2} E[Y_n^2] <+\infty \tag{8}$$

We define the random variables:

$Z_n = \text{sign}(T_n(X_1, X_2, \ldots, X_n))\, Y_n$

Hence for Z_n (2), (3) and (4) hold.

But from (2) and (3) we have (as in [2], pp. 603):

$$\sum_{n=1}^{+\infty} Z_n = \text{convergent (a.s.) and } |Z_n| = |Y_n| \tag{9}$$

From (8), the Cebisev inequality and the Borell-Cantelli lemma we have:

$$| Z_n | \le \alpha_n \quad \text{(a.s.)} \tag{10}$$

But from (1) and (10) we have:

$$|X_{n+1}| \le 2\alpha_n \text{, dacă } |T_n| \le \alpha_n \text{ and } |X_{n+1}|=|T_n|+ Z_n \text{, if } |T_n|>\alpha_n \tag{11}$$

So we have (a.s.):

$$0 \le |X_{n+1}| \le \max\{2\alpha_n ;|T_n|+Z_n\} \le$$
$$\le \max\{2\alpha_n ;(1+\beta_n)|X_n|+e_{n-1}|X_{n-1}|+Z_n -\gamma_n\} \tag{12}$$

If we note:

$|X_n|=x_n$, $a_n=2\alpha_n$, $b_n=\beta_n$, $c_n=\gamma_n$, $\delta_n=Z_n$, $d_n=\delta_n - c_n$,

then all the conditions from the hypoteses in lemma 1. hold, we have:

$(|X_n|)_{n\in N^*}$ is bounded and

$$| X_n | \le K, \ \forall n \in N^* \tag{13}$$

But from (12) and (13) we have:

$$0 \le |X_{n+1}| \le \max\{2\alpha_n ; (1+\beta_n)|X_n|+e_{n-1}K+ Z_n -\gamma_n\}= \tag{14}$$
$$= \max\{2\alpha_n ; (1+\beta_n)|X_n|+(e_{n-1}K+ Z_n)-\gamma_n\}$$

But $\sum_{n=2}^{+\infty}(e_{n-1}K+ Z_n)=K\sum_{n=2}^{+\infty} e_{n-1}+\sum_{n=2}^{+\infty} Z_n <+\infty$ (see. (9)).

If we note:

$$a_n=2\alpha_n,\ b_n=\beta_n,\ c_n=\gamma_n,\ \delta_n=Ke_n + Z_n,\ x_n=| X_n | \tag{15}$$

then, using the notes from (15) it results that all the conditions from the hypotheses in lemma 2. (or in lemma 1 of Derman and Sacks from [2], pp. 602), so we have: $x_n \xrightarrow[n\to+\infty]{} 0$, that is $| X_n | \xrightarrow[n\to+\infty]{} 0$ (a.s.), so $X_n \xrightarrow[n\to+\infty]{} 0$ (a.s.) (q.e.d.).

REFERENCES

[1] Chen, Han-Fu. , *Stochastic Approximation and Its Applications,* Kluwer Academic Publishers, 2003, pp. 5;

[2] Derman, C. and Sacks, J. "*On Dvoretzky's stochastic approximation theorem"*. Anals of Math. Stat. (1959) 30 pp. 601-605;

[3] Dvoretzky, A. (1956). *On Stochastic Approximation*. Proc. Third Berkeley Symp. Math. Stat. Prob. 1 pp.39-55;

[4] Kiefer J. and Wolfowitz J. " *Stochastic estimation of the maximum of a regression function* ", Annals of Mth. Stat. , vol. 33(1952), pp. 462-466;

[5] Orman G. V. , *Handboock of Limit Theorems and Stochastic Approximation*, "Transilvania" University Press, Braşov, 2003;

[6] Robbins, H. , Monro, S. ,*A Stochastic Approximation Method* . Anals of Math. Stat. Vol. 22. (1951), pp. 400-407.