Economy Transdisciplinarity Cognition www.ugb.ro/etc	Vol. 22, Issue 2/2019	151-156

A New Generalization of Dvoretzky's Stochastic Approximation Theorem

Neculai CRÎȘMARU
George Bacovia University, Bacau, ROMANIA
crismaru_nicolae@yahoo.com

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 personal lemmas which 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*

Introduction

Let (Ω,*K*,P) a space of probability fixed [5]. It is considered a range of functions $(T_n)_{n\in N^*}$, $T_n : R^n \to$R, T_n= measurable, $\forall$ n$\geq$1, operators which, eventually satisfy certain conditions (capable to assure the convergence of a certain type of range of random variables , which will be constructed below). Supposing we have a range of random variables $(Y_n)_{n\in N^*}$, $Y_n : \Omega \to$R, $\forall$ n$\geq$1, and with these data and, if it is considered the following random variable (called initial)
$X_1 : \Omega \to$R, it can be constructed recurrently, the following range of random variables, $(X_n)_{n\in N^*}$, $X_n : \Omega \to$R, given by the formula:

$$X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n, \ \forall\ n \geq 1.$$

Such strings of random variables frequently occur in economical applications of stochastic processes. Solving the different equations (containing random terms) from the mathematical modelling of some economic phenomena, can be made easier from the point of view of the volume of calculations performed, if we put those equations in the form:

$$X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n, \ \forall\ n \geq 1.$$

By analogy with the Robbins - Monro string name [4], we give the following definition:
The string of random variables $(X_n)_{n\in N^*}$ defined by the formula above, we call it the Dvoretzky string attached to the string $(T_n)_{n\in N^*}$ and $(Y_n)_{n\in N^*}$.

It will be shown further (by theorem .1 below) that under certain conditions this string is convergent in a certain way.

This theorem has a high degree of generality, and has benefited over many generalizations over time. We recall here the generalization from the real one-dimensional case (as it is in the form below) to the real multi-dimensional case, given by Derman and Sacks [2, 603], as well as its generalization on real Hilbert spaces, given by Venter [3, 1535-1536]. We should also mention that this Dvoretzky string is part of it from a wider class of approximation schemes, which can also be studied for the deterministic case. Also as Dvoretzky himself pointed out, Robbins-Monro and Kieffer-Wolfowitz algorithms are particular cases of this very general stochastic approximation theorem. [1, 49]

1. The Dvoretzky's Algorithm

Theorem 1. (Dvoretzky) [1, 40]

Whether $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$ three rows of real numbers so we have

(1) a_n>0, $\forall$ n$\geq$1 and $\lim_{n\to+\infty}(a_n)$=0

(2) b_n>0, $\forall$ n$\geq$1 and $\sum_{n=1}^{+\infty} b_n < +\infty$

(3) c_n>0, $\forall$ n$\geq$1 and $\sum_{n=1}^{+\infty} c_n = +\infty$

Whether $(T_n)_{n\in N^*}$, $T_n : R^n \to$R, T_n=measurable, $\forall$ n$\geq$1 so that

(4) $|T_n(t_1, t_2, \ldots, t_n) - \alpha| \leq \max\{a_n; (1+b_n)|t_n - \alpha| - c_n\}$, $\forall (t_1, t_2, \ldots, t_n) \in R^n$, $\forall$ n$\geq$1.

Considering $X_1 : \Omega \to$R, any random variable and be the string of random variables $(Y_n)_{n\in N^*}$ $Y_n : \Omega \to$R, $\forall$ n$\geq$1 and we define the string of random variables $(X_n)_{n\in N^*}$, $X_n : \Omega \to$R, given by the formula:

(5) $X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n$, $\forall$ n$\geq$1.

Suppose we have:

(6) E[X_1]<$+\infty$

(7) $\sum_{n=1}^{+\infty}$ E[$Y_n^{\ 2}$]<$+\infty$

(8) E[$Y_n \mid X_1, X_2, \ldots, X_n$]=0 (almost sure), $\forall$ n$\geq$1.

In such circumstances we have:

$\lim_{n\to+\infty}$ E[$(X_n - \alpha)^2$]=0 and (meaning, $(X_n)_{n\in N^*}$ converges on the quadratic mean to α, or (almost sure)) and

P($\{\omega \in \Omega \mid \lim_{n\to+\infty} X_n(\omega) = \alpha\}$)=1(namely $(X_n)_{n\in N^*}$ converges in probability to converge α)

The following is an extended version of the above theorem [1, 40].

Theorem 2 (Dvoretzky – extended) [1, 40]

Considering $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$ three lines of functions, so we have:

(1) $a_n : R^n \to$R, $a_n(t_1, t_2, \ldots, t_n) \geq 0$, $\forall (t_1, t_2, \ldots, t_n) \in R^n$, $\forall$ n$\geq$1, and $(a_n)_{n\in N^*}$ = a row of functions uniformly bordered on R^n and $\lim_{n\to+\infty} a_n(t_1, t_2, \ldots, t_n)$=0 uniformly, for $\forall (t_1, t_2, \ldots, t_n) \in R^n$

(2) $b_n : R^n \to$R, $b_n(t_1, t_2, \ldots, t_n) \geq 0$, $\forall (t_1, t_2, \ldots, t_n) \in R^n$, $\forall$ n$\geq$1, and b_n=measurable, $\forall$ n$\geq$1 and $\sum_{n=1}^{+\infty} b_n(t_1, t_2, \ldots, t_n)$= uniformly convergent and uniformly bounded series, $\forall (t_1, t_2, \ldots, t_n) \in R^n$

(3) $c_n : R^n \to$R thus $\sum_{n=1}^{+\infty} c_n(t_1, t_2, \ldots, t_n) = +\infty$ (uniformly, $\forall (t_1, t_2, \ldots, t_n) \in R^n$).

Considering $\alpha \in$R, and $(T_n)_{n\in N^*}$, $T_n : R^n \to$R, T_n=measurable, $\forall$ n$\geq$1 thus

(4) $|T_n(t_1, t_2, \ldots, t_n) - \alpha| \leq \max\{a_n(t_1, t_2, \ldots, t_n);$

$; (1+b_n(t_1, t_2, \ldots, t_n))|t_n - \alpha| - c_n(t_1, t_2, \ldots, t_n)\}$, $\forall\ (t_1, t_2, \ldots, t_n) \in R^n$, $\forall$ n$\geq$1

Whether $X_1 : \Omega \to$R,, any random variable and be the string of random variables $(Y_n)_{n\in N^*}$, $Y_n : \Omega \to$R, $\forall$ n$\geq$1 and we define the string of random variables $(X_n)_{n\in N^*}$, $X_n : \Omega \to$R, given by the formula

(5) $X_{n+1} = T_n(X_1, X_2, \ldots, X_n) + Y_n$, $\forall$ n$\geq$1.

Suppose we have:

(6) E[X_1]<$+\infty$

(7) $\sum_{n=1}^{+\infty}$ E[$Y_n{}^2$]<$+\infty$

(8) E[Y_n | $X_1, X_2, \ldots, X_n$]=0 (almost sure), $\forall$ n$\geq$1.

Under these circumstances:

$\lim_{n\to+\infty}$ E[$(X_n - \alpha)^2$]=0 (namely, $(X_n)_{n\in N^*}$ converges to a quadratic mean to α) and

P({$\omega \in \Omega \mid \lim_{n\to+\infty} X_n(\omega) = \alpha$ })=1 (namely $(X_n)_{n\in N^*}$ converge in probability, la α ,)

2. The Main Result. The Generalisation of Dvoretzky's Theorem

In this section we give a generalization of Dvoretzky's theorem of stochastic approximation, a generalization that is based on an auxiliary motto. But first, another helpful motto will be given with the demonstration similar to the one in. [2]

Lemma 1.

Be it the real rows $(a_n)_{n\in N^*}$, $(b_n)_{n\in N^*}$, $(c_n)_{n\in N^*}$,$(\xi_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$, thus we have:

$a_n, b_n, c_n, \xi_n \geq 0$, $\forall$ n$\in$N*, (1)

$\exists$ $n_0 \in$N* , $\forall$ n> n_0 we have $\delta_n - c_n \leq 0$, (2)

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

$\exists$ $n_1 \in$N* , (4)

$\forall$ n> n_1, $\xi_{n+1} \leq \max\{a_n ; (1+b_n)\xi_n + \delta_n - c_n\}$, $\forall$ n$\in$N*. (5)

Then $\lim_{n\to+\infty} \xi_n = 0$.

<u>Demonstration</u>**:**

For any number n$\in$N* we mark

$d_n = \delta_n - c_n$, (6)

$B_n = \prod_{i=1}^{n} (1+b_i)$ and $\theta_n = \dfrac{d_n}{B_n}$, (7)

Be it now a natural figure marked with N thus N$\geq$max{ n_0, n_1} and n$\geq$N, n=fixed . Star ting now back from n to N, we have (as in [2, 601]):

$$\xi_{n+1} \leq \{\frac{B_n}{B_{N-1}}\xi_N + B_n \sum_{j=N}^{n} \theta_j \ ; \max\{a_k \frac{B_n}{B_k} + B_n \sum_{j=k+1}^{n} \theta_j \mid N\leq k\leq n\}\} \quad (8)$$

From (2), (6), (7) and (8) we have:

$$\max\{\frac{B_n}{B_{N-1}}\xi_N+B_n\sum_{j=N}^{n}\theta_j\ ;\max\{\frac{B_n}{B_k}\mathrm{a}_k+B_n\sum_{j=k+1}^{n}\theta_j \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\}\le$$

$$\le\max\{\frac{B_n}{B_{N-1}}\xi_N+B_n\sum_{j=N}^{n}\theta_j\ ;\max\{\frac{B_n}{B_k}\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\}. \tag{9}$$

Now from (1), (2) and (3) we have

$(B_n)_{n\in N^*}$ is an increasing row to a number B=finite and $B\ge1$ (10)

But from (9) and (10) we have

$$\xi_{n+1}\le\max\{\frac{B_n}{B_{N-1}}\xi_N+B_n\sum_{j=N}^{n}\theta_j\ ;\mathrm{B}^*\max\{\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\} \tag{11}$$

Also from (10) and (11) we have

$$\xi_{n+1}\le\max\{\mathrm{B}\xi_N+\sum_{j=N}^{n}\theta_j\ ;\mathrm{B}^*\max\{\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\}=$$

$$=\mathrm{B}\max\{\xi_N+\frac{1}{B}\sum_{j=N}^{n}\theta_j\ ;\max\{\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\}, \tag{12}$$

in (3) whether it is chosen N sufficiently big, we have $\frac{1}{B}\sum_{j=N}^{n}\theta_j<0$, so

$$\mathrm{B}\max\{\xi_N+\frac{1}{B}\sum_{j=N}^{n}\theta_j\ ;\max\{\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}\}=\max\{\mathrm{a}_k \mid \mathrm{N}\le\mathrm{k}\le\mathrm{n}\}$$

So the right part of (11) can be made as small as possible (by choosing N large enough). This completes the lemma demonstration.

Observation no. 1

This lemma is still a generalization of the lemma 2.2.4. of Derman and Sacks from [2, 601]. Whether $\sum_{n=1}^{+\infty}\delta_n$=converged and $\sum_{n=1}^{+\infty}\mathrm{c}_n=+\infty$, then we get lemma.1. from [2, 601]. Or we can take the case $(\delta_n)_{n\in N^*}$=limited.

Observation no. 2

Whether rows $(\mathrm{a}_n)_{n\in N^*}$, $(\mathrm{b}_n)_{n\in N^*}$, $(\mathrm{c}_n)_{n\in N^*}$, $(\xi_n)_{n\in N^*}$, $(\delta_n)_{n\in N^*}$ are random variables that satisfy the above conditions with probability 1, then the result of the above lemma occurs with probability 1. With this lemma .1, the following generalization of Dvoretzky's theorem can be given.

Theorem 3 Considering $(\mathrm{X}_n)_{n\in N^*}$, $(\mathrm{Y}_n)_{n\in N^*}$, rows of real random variables on a space of probability (Ω, K, P) given, with X_1 arbitrary and $(\mathrm{T}_n)_{n\in N^*}$, a row of real functions measurable on R^n, (namely $\mathrm{T}_n:\ \mathrm{R}^n\to\mathrm{R}$). Marked with K_n under σ- algebra generated by functions $\{\mathrm{X}_1,\mathrm{X}_2,\ldots,\mathrm{X}_n\}$. Be it the rows of real numbers $(\alpha_n)_{n\in N^*}$, $(\beta_n)_{n\in N^*}$, $(\eta_n)_{n\in N^*}$, $(\gamma_n)_{n\in N^*}$, which are satisfying the conditions : $\alpha_n>0$, $\beta_n>0$, $\gamma_n>0$, $\eta_n>0$ and

$$\lim_{n\to+\infty}\alpha_n=0,\ \sum_{n=1}^{+\infty}\beta_n<+\infty,\ \sum_{n=1}^{+\infty}\gamma_n=\infty,\ \sum_{n=1}^{+\infty}\eta_n<+\infty, \tag{1}$$

$$\mathrm{X}_{n+1}=\mathrm{T}_n(\mathrm{X}_1,\mathrm{X}_2,\ldots,\mathrm{X}_n)+\mathrm{Y}_n \text{ with probability 1,} \tag{2}$$

Y_k is K_n-measurable for any k, $1 \le k \le n-1$ and $\forall$ $n \ge 2$, (3)

$E[Y_n | K_n]=0$ with probability 1. (4)

Be it $Z_n = \text{sign}(T_n)\, Y_n$ and it is supposed that $\sum_{n=1}^{+\infty} E[(Z_n - \eta_n)^2] < +\infty$, (5)

$\exists$ $n_0 \in N^*$, thus for any $n \in N^*$, $n \ge n_0$ we have $Z_n \le \gamma_n$ (almost sure), (6)

and $|T_n(X_1, X_2, \ldots, X_n)| \le \max\{\alpha_n; (1+\beta_n)|X_n| - \gamma_n\}$ (almost sure), $\forall$ $n \in N^*$. (7)

Then $\lim_{n \to +\infty} X_n = 0$ cu probability 1.

Demonstration: Din (3) we have the functions$(Y_k - \eta_k)$ are relatively measurable to K_n for any k $\in N^*$ thus $1 \le k \le n-1$, $n \ge 2$, and from (5) it results that Z_k is relatively measurable to K_n. Also $E[(Z_n - \eta_n)| K_n] = E[Y_n | K_n] - \eta_n = \eta_n$ and from (1) we have that

$\sum_{n=1}^{+\infty} E[(Z_n - \eta_n)| K_n] < +\infty$, with the probability 1, and $\sum_{n=1}^{+\infty} E[(Z_n - \eta_n)| K_n] \ne -\infty$.

Now from (5) and from (Lemma 3 [3, 1535], we have :

$\sum_{n=1}^{+\infty} (Z_n - \eta_n) < +\infty$ (almost sure) and $(Z_n - \eta_n) \xrightarrow[n \to +\infty]{} 0$ with the probability 1 (8)

From $\sum_{n=1}^{+\infty} \gamma_n = \infty$ and $\sum_{n=1}^{+\infty} \eta_n < +\infty$ from (1) it results that $\sum_{n=1}^{+\infty} (\eta_n - \gamma_n) = -\infty$

(*)

For any $n \in N^*$ be it the sets

$A_n = \{\omega \in \Omega \,|\, |T_n(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))| \le \alpha_n\}$ and $\overline{A}_n = \Omega - A_n$.

Be it a $\omega \in \overline{A}_n$, then $|T_n(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))| > \alpha_n$, $\forall$ $n \in N^*$.

From (2) it results that for any $\omega \in A_n$ we have:

$|X_{n+1}(\omega)| \le |T_n(X_1, X_2, \ldots, X_n)| + |Y_n(X_1, X_2, \ldots, X_n)| \le \alpha_n + |Y_n(\omega)|$ with the probability 1, and for any $\omega \in \Omega$ we have (also see T .1., from [2, 603]):

$|X_{n+1}(\omega)| \le |T_n(X_1, X_2, \ldots, X_n)| + \text{sign}(T_n(\omega))|Y_n(\omega)| = |T_n(\omega)| + Z_n(\omega)$.

So for any $\omega \in \Omega$ we have:

$|X_{n+1}(\omega)| \le \max\{\alpha_n + |Y_n(\omega)|; |T_n(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))| + Z_n(\omega).\}$, $\forall$ $n \in N^*$. (9)

From (7) and (9) we have:

$|X_{n+1}(\omega)| \le \max\{\alpha_n + |Y_n(\omega)|; \max\{\alpha_n; (1+\beta_n)|X_n(\omega)| - \gamma_n\} + Z_n(\omega)\} =$

$= \max\{\max\{\alpha_n + |Y_n(\omega)|; \alpha_n + Z_n(\omega)\}; (1+\beta_n)|X_n(\omega)| - \gamma_n + Z_n(\omega)\}$ (10)

But $|Y_n(\omega)| = |Z_n(\omega)|$, $\forall$ $\omega \in \Omega$, so we have

$\max\{\alpha_n + |Y_n(\omega)|; \alpha_n + Z_n(\omega)\} = \alpha_n + |Y_n(\omega)|$ (11)

From (10) and (11) we have

$|X_{n+1}(\omega)| \le \max\{\alpha_n + |Y_n(\omega)|; (1+\beta_n)|X_n(\omega)| + Z_n(\omega) - \gamma_n\}$ (12)

From (8) we can write

$0 \le \| Z_n | - \eta_n | \le | Z_n - \eta_n | \xrightarrow[n \to +\infty]{} 0$ with probability 1.

Conclusions

So we have

$|Y_n| = | Z_n | \le (| Z_n - \eta_n | + \eta_n) \xrightarrow[n \to +\infty]{} 0$ so from (8) and from (1) we have

$|Y_n | \xrightarrow[n \to +\infty]{} 0 |$ with probability 1. (12')

So whether in the relation (12) we do the marks:

$a_n = \alpha_n + |Y_n(\omega)| \ge 0$ (almost sure), $b_n = \beta_n$, $\xi_n = |X_n|$, $\delta_n = Z_n$, and $c_n = \gamma_n$ (13)

we have:

$\xi_{n+1} \le \max\{a_n ; (1+b_n)\xi_n + \delta_n - c_n\}, \forall\ n \in N^*$, namely just the relation (5) from lemma 1. above.

We also have the relations:

$Z_n - \gamma_n = (Z_n - \eta_n) + (\eta_n - \gamma_n)$, so from the relation $\sum_{n=1}^{+\infty} (Z_n - \eta_n) < +\infty$ (v. (8)) and from (*) we have

$\sum_{n=1}^{+\infty} (Z_n - \gamma_n) = -\infty$ with probability 1.

Finally, from the lemma .1 above we have that $X_n \xrightarrow[n \to +\infty]{} 0$

With probability 1, and thus the theorem is demonstrated.

References

[1] Dvoretzky A., (1956), *On Stochastic Approximation,* Proc. Third Berkeley Symp. Math. Stat. Prob. 1

[2] Derman C., Sacks J. (1959), *On Dvoretzky's stochastic approximation theorem*, Anals of Math. Stat. 30

[3] Venter J.H., (1966), *On Dvoretzky Approximation Theorems*, Ann. Math. Stat. 37

[4] Chen Han-Fu, (2002), *Stochastic Approximation and Its Application*, Kluver Academic Publisher, Springer

[5] Orman G.V., (2003), *Handboock of Limit Theorems and Stochastic Approximation*, Transilvania University Press, Brasov