

**NON-STANDARD LIMIT THEOREMS  
FOR URN MODELS AND STOCHASTIC  
APPROXIMATION PROCEDURES**

Yu. Kaniovski

*International Institute for Applied Systems Analysis  
Laxenburg, Austria*

G. Pflug

*Institute of Statistics and Computer Science  
University of Vienna, Austria*

RR-95-8

November 1995

Reprinted from *Stochastic Models*, Volume 11, Number 1, pp. 79–102, 1995.

**INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS  
Laxenburg, Austria**

*Research Reports*, which record research conducted at IIASA, are independently reviewed before publication. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

---

Reprinted with permission from *Stochastic Models*, Volume 11, Number 1, 1995.  
Copyright ©1995 Marcel Dekker, Inc.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the copyright holder.

---

Printed by Novographic, Vienna, Austria.

# Foreword

Stochastic approximation procedures are important general models of learning and evolution. The analysis of their limit properties helps to understand the nature of patterns that may emerge from learning. It also provides us with tools for estimating the speed of learning and the errors of the outcomes. This paper considers a particularly difficult and important class of stochastic approximation procedures in which the underlying regression functions can be nonsmooth or discontinuous. Such situations are typical for many practical situations, where the rules of behavior are different in various regions of the decision space. Non-standard limit distributions are derived, including the case of many potential equilibria.

*Andrzej Ruszczyński*

Leader

Optimization Under Uncertainty Project



**NON-STANDARD LIMIT THEOREMS FOR URN MODELS  
AND STOCHASTIC APPROXIMATION PROCEDURES**

**Yu. Kaniovski**

International Institute for Applied Systems Analysis  
Laxenburg, Austria, A-2361

and

**G. Pflug**

Institute of Statistics and Computer Science  
University of Vienna, Universitätstraße 5, Vienna, Austria, A-1010

and

International Institute for Applied Systems Analysis

**Key Words:** generalized urn scheme, conditional limit theorems, stochastic approximation.

**ABSTRACT**

The adaptive processes of growth modeled by a generalized urn scheme have proved to be an efficient tool for the analysis of complex phenomena in economics, biology and physical chemistry. They demonstrate non-ergodic limit behavior with multiple limit states. There are two major sources of complex feedbacks governing these processes: nonlinearity (even local, which is caused by nondifferentiability of the functions driving them) and multiplicity of limit states stipulated by the nonlinearity.

We suggest an analytical approach for studying some of the patterns of complex limit behavior. The approach is based on conditional limit theorems. The corresponding limits are, in general, not infinitely divisible. We show that convergence rates could be different for different limit states. The rates depend upon the smoothness (in neighborhoods of the limit states) of the functions governing the processes.

Since the mathematical machinery allows us to treat a quite general class of recursive stochastic discrete-time processes, we also derive corresponding limit theorems for stochastic approximation procedures. The theorems yield new insight into the limit behavior of stochastic approximation procedures in the case of nondifferentiable regression functions with multiple roots.

## 1 MOTIVATION AND FORMULATION OF THE PROBLEM

As a conceptual example where the generalized urn scheme can be applied consider the following model of technological dynamics. Suppose that two firms, producing technologies  $A$  and  $B$ , compete for a market of infinite capacity. To achieve market domination, they use certain strategies. Assume that prices of the technologies depend only on their market shares. Then the following strategy was introduced in [3]: until a certain level of market share (usually greater than  $1/2$ ) the price is a decreasing function, above that level it is an increasing function. Let us consider the simplest i.e. piecewise linear case of this policy. It is graphically represented in Figure 1. Let  $x_A$ , respectively  $x_B$ , be the proportion of the adopters of technology  $A$  and  $B$ , respectively ( $x_A + x_B = 1$ ).  $Pr_A(x_A)$  designates the dependence of the price of the technology  $A$  on  $x_A$ . Also  $Pr_B(x_A)$  designates the dependence of the price of the technology  $B$  on  $x_A$ . By  $x_A^*$  and  $x_B^*$  we designate the levels of market shares where the firms switch from falling- to rising-price rules. Hence the dependence of the price of the  $A$  ( $B$ ) technology on its proportion of the market  $x_A$  ( $x_B$ ) is given by four parameters:  $Pr_A(0)$ ,  $x_A^*$ ,  $Pr_A(x_A^*)$ ,  $Pr_A(1)$  ( $Pr_B(1)$ ,  $x_B^*$ ,  $Pr_B(1 - x_B^*)$ ,  $Pr_B(0)$ ). Note, that we allow also that  $Pr_A(1) \leq Pr_A(x_A^*)$  ( $Pr_B(0) \leq Pr_B(1 - x_B^*)$ ) and that  $x_A^* = 1$  (or  $x_B^* = 0$ ): in this case, firm  $A$  ( $B$ ) still reduces the price on its product as its proportion of the market increases. Note that the model can account for all possible combinations of behaviors of the firms: both lower (increase) prices as their market shares increase, one increases and the other lowers price as the corresponding share increases, one increases (lowers) price while the other follows the above general strategy. These special cases can be obtained by changing relations between  $Pr_A(0)$ ,  $Pr_A(x_A^*)$ ,  $Pr_A(1)$  ( $Pr_B(1)$ ,  $Pr_B(1 - x_B^*)$ ,  $Pr_B(0)$ ).

It is natural to suppose that in the case when the quality of the technologies is approximately the same and potential consumers know about it, the technology which is cheaper has more chances to be sold, i.e. the

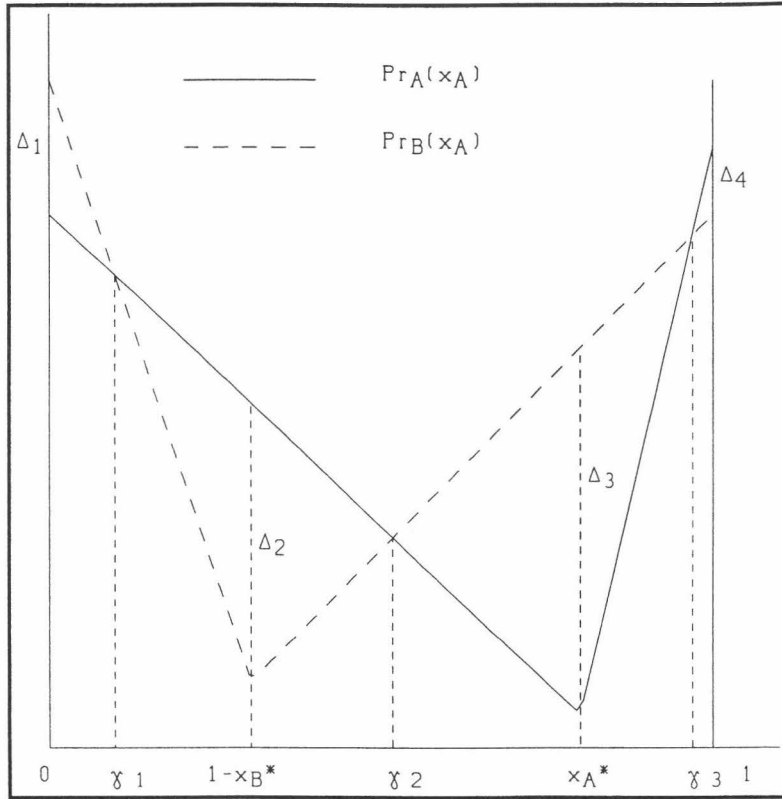


FIG. 1. Dependence of prices of  $A$  and  $B$  on the market share of  $A$ .

technology  $A$  is bought if  $Pr_A(x_A) - Pr_B(x_A) < 0$ . But if the prices differ only slightly or consumers have some specific preferences (which can be characterized only statistically or on average), then sometimes the more expensive technology will be adopted. We formalize this situation in the following way: technology  $A$  is bought if  $Pr_A(x_A) - Pr_B(x_A) + \xi < 0$ , where  $\xi$  is a random variable. Then the probability  $f(x_A)$  of choosing the technology  $A$ , as a function of  $x_A$ , equals  $P\{\xi < Pr_B(x_A) - Pr_A(x_A)\}$ . To avoid unnecessary sophistications of the model, we shall assume at the beginning that  $\xi$  has a uniform distribution on  $[-\alpha, \alpha]$ . In this case the probability of choosing  $A$  as a function of  $x_A$  has the form

$$f(x_A) = \begin{cases} 1 & \text{if } Pr_B(x_A) - Pr_A(x_A) \geq \alpha, \\ 0 & \text{if } Pr_B(x_A) - Pr_A(x_A) \leq -\alpha, \\ \frac{Pr_B(x_A) - Pr_A(x_A) + \alpha}{2\alpha} & \text{if } Pr_B(x_A) - Pr_A(x_A) \in (-\alpha, \alpha). \end{cases}$$

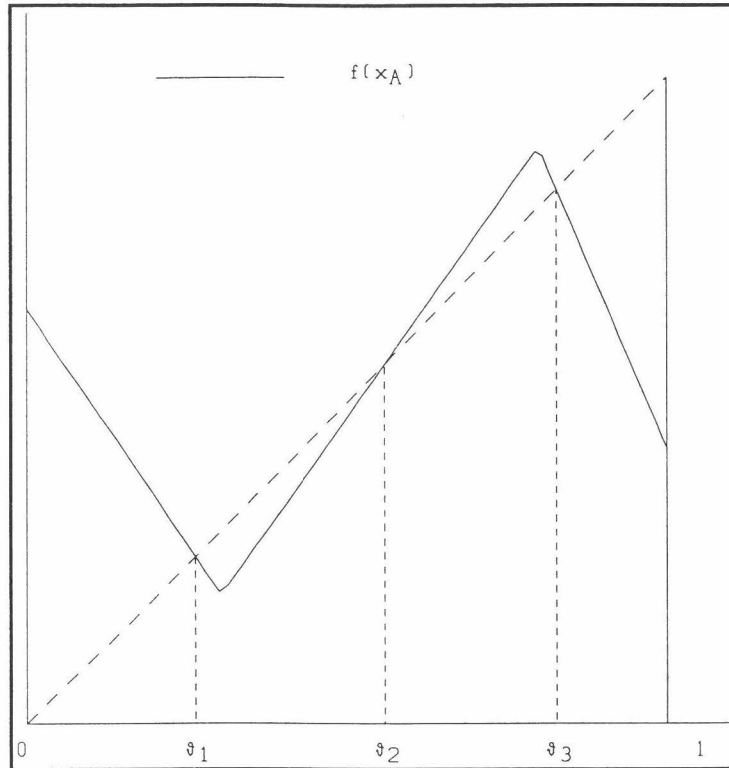


FIG. 2 Probability of choosing  $A$  depending on its market share.

For  $\alpha > \max_{i=1,2,3,4} \Delta_i$  (recall that  $\Delta_1 = Pr_B(0) - Pr_A(0)$ ,  $\Delta_2 = Pr_A(1 - x_B^*) - Pr_B(1 - x_B^*)$ ,  $\Delta_3 = Pr_B(x_A^*) - Pr_A(x_A^*)$ ,  $\Delta_4 = Pr_A(1) - Pr_B(1)$ ) this is represented graphically in Figure 2.

Now we put this problem into the framework of the generalized urn scheme and interpret the market as an urn, a unit of  $A(B)$  as a white (black) ball and the addition of balls into the urn as adoption of the technologies. By a generalized urn scheme we mean the following.

Think of an urn of infinite capacity with balls of two colors, say black and white. Starting from  $w_1 \geq 0$  white balls and  $b_1 \geq 0$  black ones ( $\gamma_1 = w_1 + b_1 \geq 1$ ), a ball is added into the urn at time instants  $t = 1, 2, \dots$ . It will be black with probability  $f(X_t)$  and white with probability  $1 - f(X_t)$ . Here  $f(\cdot)$  stands for a function which maps  $[0, 1]$  into itself and  $X_t$  is the proportion of black balls already present in the urn at time  $t$  before the  $(\gamma_1 + t)$ -th ball is added. The function  $f(\cdot)$  is called the urn function



for this generalized urn scheme (see, for example, [5]). Consider some probability space  $(\Omega, \mathcal{F}, P)$ . For  $t \geq 1$  and  $x \in \mathbf{Q} \cap [0, 1]$ , let

$$\xi_t(x) = \begin{cases} 1 & \text{with probability } f(x), \\ 0 & \text{with probability } 1 - f(x), \end{cases}$$

be (independent in  $t$ ) random variables defined on  $(\Omega, \mathcal{F}, P)$ . Here  $\mathbf{Q}$  stands for the set of rational numbers. Then the stochastic process  $\{X_t\}$  follows the dynamics

$$\begin{aligned} X_{t+1} &= X_t + \frac{1}{\gamma_1 + t} [\xi_t(X_t) - X_t] = X_t + \frac{1}{\gamma_1 + t} [f(X_t) - X_t] + \\ &\frac{1}{\gamma_1 + t} \eta_t(X_t), \quad t \geq 1, \quad X_1 = b_1/\gamma_1, \end{aligned} \quad (1.1)$$

where  $\eta_t(x) = \xi_t(x) - f(x)$ . The recurrent process (1.1) can be considered as a particular case of the well known Robbins–Monro stochastic approximation procedure.

Suppose that we are interested in finding a root of a continuous regression function  $g(x)$ ,  $x \in \mathbf{R}$ , the set of real numbers. We cannot observe  $g(\cdot)$  itself, but only

$$y(s, x) = g(x) + z_s(x), \quad s \geq 1, \quad x \in \mathbf{R}, \quad (1.2)$$

independent (in  $s$ ) observations of  $g(\cdot)$  together with a random noise  $z_s(\cdot)$ . It is assumed that the noise has zero mean, i.e.  $Ez_s(x) = 0$ , finite variance, i.e.  $Ez_s^2(x) = \sigma_s^2(x) < \infty$ , and there is a mapping  $n(s, \cdot, \cdot)$ , measurable with respect to  $\mathcal{B}_2$ , such that  $z_s(x) = n(s, x, \xi_s)$ . Here  $\mathcal{B}_2$  stands for the  $\sigma$ -algebra of Borel sets in the real plane  $\mathbf{R}^2$ . Also  $\xi_s, s \geq 1$ , is a sequence of independent random variables on  $(\Omega, \mathcal{F}, P)$ . Then the Robbins–Monro procedure gives successive approximations  $X_n, n \geq 1$ , to the root  $\Theta$  of  $g(\cdot)$  in the following form:

$$X_{n+1} = X_n - \gamma_n y(n, X_n), \quad X_1 = \text{const}, \quad (1.3)$$

where  $\gamma_n$  stands for the step-sizes, i.e. deterministic positive numbers such that

$$\sum_{n \geq 1} \gamma_n = \infty, \quad \sum_{n \geq 1} \gamma_n^2 < \infty.$$

Many results are known for the case of a unique root  $\Theta$ . For  $\gamma_n = an^{-1}$  asymptotic normality of  $\sqrt{n}(X_n - \Theta)$  was proved (see [2, 4, 11, 17]) for locally linear  $g(\cdot)$  at  $\Theta$ , i.e.  $g(x) = g'(\Theta)(x - \Theta) + o(|x - \Theta|)$  as

$x \rightarrow \Theta$ , and  $ag'(\Theta) > 1/2$ . Also  $\sqrt{\frac{n}{\ln n}}(X_n - \Theta)$  is asymptotically normal if  $\alpha g'(\Theta) = 1/2$  [10]. But for the simplest case, when  $g(\cdot)$  is not locally linear at  $\Theta$ , i.e., as  $x \rightarrow \Theta$

$$g(x) = \begin{cases} \alpha_1(x - \Theta) + o_1(x - \Theta), & x \geq \Theta, \\ \alpha_2(\Theta - x) + o_2(\Theta - x), & x < \Theta, \end{cases} \quad (1.4)$$

$\sqrt{n}(X_n - \Theta)$  converges weakly for  $2a \min(\alpha_1, \alpha_2) > 1$  to a non-Gaussian limit distribution [7, 9]. By  $o_i(y)$  we designate a function of  $y$  such that

$$\lim_{y \rightarrow 0} \frac{|o_i(y)|}{y} = 0.$$

In this paper we study the limit behavior of properly normalized deviations of  $X_n$  from  $\Theta$  for the case when, as  $x \rightarrow \Theta$ ,

$$g(x) = \begin{cases} \alpha_1(x - \Theta)^\gamma + o_1((x - \Theta)^\gamma), & x \geq \Theta, \\ -\alpha_2(\Theta - x)^\gamma + o_2((\Theta - x)^\gamma), & x < \Theta, \end{cases} \quad (1.5)$$

for some  $\alpha_1 > 0$ ,  $\alpha_2 > 0$  and  $\gamma \in (\frac{1}{2}, 1)$ . Also, we consider the cases where random processes of the form (1.3) demonstrate complex limit behavior caused by both the nonlinearity of the form (1.4) or (1.5) of functions involved and the multiplicity of their roots. Comparing (1.1) and (1.3), one sees that (1.1) is a special case of (1.3) with  $g(x) = f(x) - x$  and  $z_t(x) = \eta_t(x)$ .

Returning to the above model of technological dynamics we observe that the roots  $\theta_1$  and  $\theta_3$  satisfy condition (1.4). Consequently the limit theorems we are going to derive will characterize the convergence rates in time of the proportion of  $A$  on the market to  $\theta_1$  and  $\theta_3$  (it turns out that the probability that  $X_t$  has  $\theta_2$  as a limit point is zero [5]). In the next section we prove limit theorems for the random process (1.3).

## 2 LIMIT THEOREMS FOR THE ROBBINS – MONRO PROCEDURE IN NON-STANDARD SITUATIONS

We start with an auxiliary lemma (see [9], Lemma 2.1).

**Lemma 2.1.** If  $\{y_n\}$  is a sequence of real numbers such that

$$|y_{n+1}| \leq |y_n| (1 - b_n) + c_n,$$

where  $\sum_{n \geq 1} b_n = \infty$ ,  $b_n \rightarrow 0$ ,  $c_n \geq 0$ ,  $b_n > 0$ . Then  $|y_n| = o(1)$  or  $|y_n| = O(1)$  depending upon whether  $c_n = o(b_n)$  or  $c_n = O(b_n)$ .

Here, as usual,  $O(y)$  means that

$$\limsup_{y \rightarrow 0} \frac{|O(y)|}{y} < \infty.$$

We study the algorithm (1.3) with (the unique root)  $\Theta = 0$  and  $\gamma_n = an^{-1}$ ,  $a > 0$ . For a real valued function  $h(\cdot)$  we set  $\|h\| = \sup_x |h(x)|$ .

**Theorem 2.1.** Assume that

1.  $xg(x) \geq \alpha_0 x^2$  for an  $\alpha_0 > 0$ ;
2.  $|g(x)| \leq A|x| + B$  for some constants  $A, B > 0$ ;
3.  $g(\cdot)$  has the form (1.5) with  $\Theta = 0$  and  $|o_i(y)| = O(y^{1+\nu/\gamma})$  for  $\nu > \frac{1-\gamma}{1+\gamma}$ ;
4.  $E|z_s(x) - z_s(0)|^2 \leq k|x|^\nu$ , where  $\nu > \frac{(1-\gamma)^2}{1+\gamma}$ ;
5. for some  $\sigma > 0$  and  $\kappa > \frac{1-\gamma}{2(1+\gamma)}$  one has  $\lim_{s \rightarrow \infty} |Ez_s(0)^2 - \sigma^2|s^\kappa = 0$ ;
6.  $\sup_s E|z_s(0)|^\mu < \infty$  for some  $\mu > 2 + \frac{1-\gamma}{2(1+\gamma)(2\gamma-1)}$ .

Then for  $\frac{1}{2} < \gamma < 1$

$$n^{\frac{1}{1+\gamma}} X_n \xrightarrow[n \rightarrow \infty]{w} X,$$

where  $X$  has the density

$$f(x) = C \begin{cases} \exp \left\{ -\frac{2\alpha_1}{a\sigma^2} \frac{x^{1+\gamma}}{1+\gamma} \right\}, & x \geq 0, \\ \exp \left\{ -\frac{2\alpha_2}{a\sigma^2} \frac{(-x)^{1+\gamma}}{1+\gamma} \right\}, & x < 0. \end{cases}$$

Here  $C$  stands for a normalizing constant.

The first two conditions here are standard. They can be easily relaxed (see Remark 2.1). The third condition specifies the nonlinearity we are dealing with. If we formally put here  $\gamma = 1$ , which corresponds in the simplest situation to the conventional case (of a differentiable regression function), this condition, requiring a certain rate of decrease for the small term, proves to be somewhat stronger than usual (see (1.4)). We cannot relax it because of purely technical reasons. Conditions 4, 5 and 6 postulate that the noise term can be approximated by independent centered random variables with equal variances and having some higher moments. If, as in the conventional case, we formally set here  $\gamma = 1$ , this requirement of higher moments converts in to Liapounov's condition, which is a little stronger than the conventional Lindeberg's one. Now let us prove the theorem.

**Proof.** Set  $\beta = \frac{1}{1+\gamma}$ , i.e.  $\gamma\beta = 1 - \beta$ , and let  $U_n = n^\beta X_n$ . Then

$$U_{n+1} = U_n - a_n^2 h_n(U_n) + a_n z_n(n^{-\beta} U_n), \quad (2.1)$$

where

$$\begin{aligned} a_n &= an^{-1}(n+1)^\beta = an^{\beta-1}(1+\varepsilon_n), \quad \varepsilon_n = O(n^{-1}), \quad h_n(u) = \\ & -a_n^{-2}[(1+\frac{1}{n})^\beta - 1]u + a_n^{-1}g(n^{-\beta}u) = -a^{-2}\beta n^{1-2\beta}(1+\eta_n)u + \\ & a^{-1}n^{1-\beta}(1+\varepsilon_n)^{-1}g(n^{-\beta}u), \quad \eta_n = O(n^{-1}). \end{aligned}$$

We will replace the functions  $h_n(\cdot)$  by simpler functions  $h_n^*(\cdot)$  and show that this has no effect on the asymptotic distribution.

Set  $k_n = n^{\beta+\delta-1/2}$ ,  $n \geq 1$ , with  $0 < \delta < \frac{1}{2}$  to be fixed later. Then  $k_n \rightarrow \infty$  since  $\beta > 1/2$ . Our assumptions imply that  $n^{1/2-\delta}X_n \xrightarrow{a.s.} 0$  for every  $\delta > 0$  (c.f. [4], Lemma 2.3), and consequently  $k_n^{-1}U_n \xrightarrow{a.s.} 0$ .

We construct functions  $h_n^*(\cdot)$  with the following properties

$$h_n^*(u)\operatorname{sgn}u \geq 0, \quad \int_0^\infty h_n^*(u)du = \int_0^{-\infty} h_n^*(u)du = \infty, \quad (2.2)$$

$$\|h_n^*\| \|h_n^{*'}\| = o(a_n^{-1}), \quad (2.3)$$

$$\sup_{|u| \leq k_n} |h_n^{*''}(u)| = O(n^{\varepsilon(2-\gamma)}), \quad \sup_{|u| \leq k_n} |h_n^{*'''}(u)| = O(n^{\varepsilon(3-\gamma)}), \quad (2.4)$$

$$a_n^2 \inf_{|u| \leq k_n} |h_n^{*'}(u)| \geq cn^{-1/2\beta}, \quad (2.5)$$

$$a_n^2 \sup_{|u| \leq k_n} |h_n(u) - h_n^*(u)| = o(n^{-1/2\beta}), \quad (2.6)$$

where  $\varepsilon > 0$  will be chosen later. To this end, let  $h_n^*(\cdot)$  be a smoothed modification of

$$\bar{h}_n(u) = \begin{cases} C_n^{(1)}, & u > n^{\beta-\varepsilon}, \\ \frac{\alpha_1}{a}u^\gamma, & n^{-\varepsilon} < u \leq n^{\beta-\varepsilon}, \\ C_n^{(2)}u, & 0 \leq u \leq n^{-\varepsilon}, \\ -C_n^{(3)}u, & -n^{-\varepsilon} \leq u < 0, \\ -\frac{\alpha_2}{a}|u|^\gamma, & -n^{\beta-\varepsilon} \leq u < -n^{-\varepsilon}, \\ -C_n^{(4)}, & u < -n^{\beta-\varepsilon}. \end{cases}$$

Here  $C_n^{(i)}$  are chosen in order to make  $\bar{h}_n(\cdot)$  continuous and the smoothing is done to make  $h_n^*(\cdot)$  three times differentiable. Relations (2.2) and (2.4) are obvious. Also (2.3) follows by

$$\|h_n^*\| \|h_n^{*'}\| = O(n^{\gamma(\beta-\varepsilon)})O(n^{\varepsilon(1-\gamma)}) = O(n^{1-\beta-\varepsilon(2\gamma-1)}) = o(a_n^{-1})$$

and (2.5) follows from the fact that there is a constant  $c_1$  with

$$\inf_{\|u\| \leq k_n} h_n^{*'}(u) \geq c_1 n^{-(1-\gamma)^2 \beta/2}.$$

In order to show (2.6) notice that for  $0 \leq u \leq k_n$

$$\begin{aligned} |n^{1-\beta} g(n^{-\beta} u) - \alpha_1 |u^\gamma| | &= n^{1-\beta} O((n^{-\beta} u)^{\gamma+\nu}) = \\ O(n^{1-\beta+(\gamma+\nu)(\delta-1/2)}) &= O(n^{-\xi}). \end{aligned} \quad (2.7)$$

Since, by assumption,  $\nu > 2\beta - 1$  we may choose  $\delta$  so small that  $-\xi = 1 + \beta - (\gamma + \nu)(\delta - 1/2) < 2 - 2\beta - 1/2\beta$ , i.e.

$$n^{-\xi} a_n^2 = o(n^{-1/2\beta}).$$

The same is true for  $-k_n \leq u \leq 0$ . Consider now the recursion

$$W_{n+1} = W_n - a_n^2 h_n^*(W_n) + a_n z_n, n \geq \mathcal{N}, \quad (2.8)$$

with  $W_{\mathcal{N}}$  arbitrary (but it does not depend on  $z_n(\cdot), n \geq 1$ ) and  $\mathcal{N}$  some positive integer.

Let  $\tau$  be the stopping time

$$\tau = \inf\{n \geq \mathcal{N} : \max(|U_n|, |W_n|) > k_n\}.$$

Since  $k_n^{-1} U_n \rightarrow 0$  and  $k_n^{-1} W_n \rightarrow 0$  a.s.,  $P\{\tau = \infty\}$  can be made arbitrarily close to 1 by choosing  $\mathcal{N}$  large. On the event  $\{\tau = \infty\}$ , using the bounds (2.5) and (2.6), we get

$$\begin{aligned} |U_{n+1} - W_{n+1}| &\leq |U_n - W_n - a_n^2 [h_n(U_n) - h_n^*(U_n)]| + \\ a_n^2 |h_n^*(U_n) - h_n^*(W_n)| &\leq |U_n - W_n| (1 - cn^{-1/2\beta}) + o(n^{-1/2\beta}). \end{aligned}$$

By  $\sum_{n \geq 1} n^{-1/2\beta} = \infty$  and the auxiliary lemma we get  $|U_n - W_n| \rightarrow 0$  on  $\{\tau = \infty\}$ . It is therefore sufficient to consider the asymptotic behavior of  $W_n$ .

In the next step we show that without affecting the asymptotic distribution, the recursion (2.7) can be replaced by the following

$$V_{n+1} = V_n - a_n^2 h_n^*(V_n) + a_n z_n^*, n \geq \mathcal{N}, V_{\mathcal{N}} \text{ arbitrary}, \quad (2.9)$$

where  $z_n^* = z_n(0)$  and  $V_{\mathcal{N}}$  does not depend on  $z_n^*, n \geq \mathcal{N}$ . Introduce  $\tau^* = \inf\{n \geq \mathcal{N} : \max(|V_n|, |W_n|) > k_n\}$ .

Using condition 4, one has

$$E[(z_n - z_n^*)^2 \chi_{\{\tau^* > n\}}] \leq kE[(n^\beta |W_n|)^\nu \chi_{\{\tau^* > n\}}] \leq kn^{(\delta-1/2)\nu}. \quad (2.10)$$

Here  $\chi_A$  stands for the indicator function of the event  $A$ . Also there are constants  $c_2$  and  $c_3$  such that

$$|h_n^*(u) - h_n^*(v)| \leq c_2 + c_3|u - v|. \quad (2.11)$$

If  $\delta$  is so small that  $(\delta - 1/2)\nu < -(1 - \gamma)^2\beta/2$ , then from (2.5), (2.9) and (2.10) we have

$$\begin{aligned} E(W_{n+1} - V_{n+1})^2 \chi_{\{\tau^* > n+1\}} &\leq E(W_n - V_n)^2 \chi_{\{\tau^* > n\}} - \\ &2a_n^2 E(W_n - V_n)[h_n^*(W_n) - h_n^*(V_n)] \chi_{\{\tau^* > n\}} + \\ &a_n^2 E(z_n - z_n^*)^2 \chi_{\{\tau^* > n\}} + a_n^4 E[h_n^*(W_n) - h_n^*(V_n)]^2 \chi_{\{\tau^* > n\}} \leq \\ &(1 - c_4 n^{-1/2\beta}) E(W_n - V_n)^2 \chi_{\{\tau^* > n\}} + o(n^{-1/2\beta}). \end{aligned}$$

Hence due to Lemma 2.1

$$E(W_n - V_n)^2 \chi_{\{\tau^* > n\}} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

which implies that  $W_n$  and  $V_n$  have the same limit distribution.

Due to condition 5, arguments identical to the ones given above show that the recursion of the form (2.8) with  $z_n^*$  substituted by

$$\sigma[Ez_n(0)^2]^{-1/2} z_n(0)$$

has the same limit distribution. Consequently, we can suppose that  $E(z_n^*)^2 = \sigma^2, n \geq \mathcal{N}$ .

We will now replace  $z_n^*$  by truncated vectors  $z_n^{**}, n \geq \mathcal{N}$ . Consider

$$z_n^{**} = \begin{cases} z_n^* & \text{if } |z_n^*| \leq n^b, \\ 0 & \text{otherwise.} \end{cases}$$

Here  $b$  satisfies the inequality

$$b > \frac{1 - \gamma}{2(1 + \gamma)^2(\mu - 2)}. \quad (2.12)$$

By Markov's inequality

$$|Ez_n^{**}| \leq n^{b(1-\mu)} E|z_n^*|^\mu \quad (2.13)$$

and

$$|\mathbb{E}z_n^{**}|^2 \leq n^{b(2-\mu)}\mathbb{E}|z_n^*|^\mu.$$

Consequently for  $z_n^{***} = z_n^{**} - \mathbb{E}z_n^{**}$  one has  $\mathbb{E}z_n^{***} = 0$  and  $\mathbb{E}(z_n^* - z_n^{***})^2 = O(n^{b(2-\mu)})$ .

Due to (2.11),

$$a_n^2 O(n^{b(2-\mu)}) = o(n^{-1/2\beta})$$

and we can replace  $z_n^*$  by  $z_n^{***}$  without changing the asymptotic distribution (the arguments are the same as above).

Also we can substitute  $z_n^{***}$  by  $\tilde{z}_n = \sigma(\mathbb{E}(z_n^{***})^2)^{-1/2}z_n^{***}$  without affecting the limit behavior. This can be done by the same reasoning since by Markov's inequality

$$|\mathbb{E}(z_n^{**})^2 - \sigma^2| \leq n^{b(2-\mu)}\mathbb{E}|z_n^*|^\mu$$

and

$$\text{Var } z_n^{***} = \text{Var } z_n^{**} = \mathbb{E}(z_n^{**})^2 - (\mathbb{E}z_n^{**})^2$$

which, together with (2.12) implies that  $|\text{Var } z_n^{***} - \sigma^2| = O(n^{b(2-\mu)})$ .

From now on we consider the recursion

$$V_{n+1} = V_n - a_n^2 h_n^*(V_n) - a_n \tilde{z}_n, n \geq \mathcal{N},$$

where  $V_{\mathcal{N}}$  -arbitrary (but it does not depend on  $\tilde{z}_n, n \geq \mathcal{N}$ ). Notice that for large enough  $n$

$$|\tilde{z}_n| \leq 2n^b \text{ a.s.}, \mathbb{E}\tilde{z}_n = 0, \text{Var } \tilde{z}_n = \sigma^2.$$

Consider the function  $H_n(x) = x - a_n^2 h_n^*(x)$ . Since  $\sup_x |H_n'(x) - 1| < 1/2$  for sufficiently large  $n$ , we have by (2.3)

$$\begin{aligned} |H_n^{-1}(x) - [x + a_n^2 h_n^*(x)]| &\leq 2|x - H_n(x + a_n^2 h_n^*(x))| \leq \\ a_n^4 \|h_n^*\| \|h_n^*\| &= O(a_n^3). \end{aligned} \quad (2.14)$$

If  $F_n(\cdot)$  stands for the distribution function of  $V_n$ , then  $V_{n+1}$  is distributed according to

$$T_n(F_n)(x) = \int F_n(H_n^{-1}(z)) dG_n\left(\frac{x-z}{a_n}\right),$$

where  $G_n(\cdot)$  is the distribution function of  $\tilde{z}_n$ . Let  $F_n^*(\cdot)$  be the distribution with density

$$f_n^*(x) = C_n \exp\left[-\frac{2}{\sigma^2} \int_{-\infty}^x h_n^*(u) du\right],$$

where  $C_n$  is a normalizing constant. We show that  $T_n(F_n^*)(\cdot)$  is close to  $F_n^*(\cdot)$ , i.e.  $F_n^*(\cdot)$  is nearly a stationary distribution. We know from (2.13) that

$$\sup_x |F_n^*(H_n^{-1}(x)) - F_n^*(x + a_n^2 h_n^*(x))| = O(a_n^3).$$

By a Taylor expansion up to the order three, we get ( $\tilde{x}$  is some interpolation point)

$$\begin{aligned} T_n(F_n^*)(x) &= \int F_n^*(z + a_n^2 h_n^*(z)) dG_n\left(\frac{x-z}{a_n}\right) + O(a_n^3) = \\ &F_n^*(x + a_n^2 h_n^*(x)) + \int (x-z) \frac{\partial}{\partial x} [F_n^*(x + a_n^2 h_n^*(x))] dG_n\left(\frac{x-z}{a_n}\right) + \\ &\frac{1}{2} \int (x-z)^2 \frac{\partial^2}{\partial x^2} [F_n^*(x + a_n^2 h_n^*(x))] dG_n\left(\frac{x-z}{a_n}\right) + \\ &\frac{1}{6} \int (x-z)^3 \frac{\partial^3}{\partial x^3} [F_n^*(\tilde{x} + a_n^2 h_n^*(\tilde{x}))] dG_n\left(\frac{x-z}{a_n}\right) + O(a_n^3) = \end{aligned} \quad (2.15)$$

$$\begin{aligned} &F_n^*(x + a_n^2 h_n^*(x)) + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} [F_n^*(x + a_n^2 h_n^*(x))] + \\ &O(a_n^3 n^b n^{\varepsilon(3-\gamma)}) + O(a_n^3) = \end{aligned}$$

$$F_n^*(x) + a_n^2 f_n^*(x) h_n^*(x) + a_n^2 \frac{\sigma^2}{2} f_n^{*'}(x) + O(a_n^3 n^{b+\varepsilon(3-\gamma)}).$$

Due to condition 6 we can choose  $b$  satisfying (2.11) such that

$$b < \frac{2\gamma - 1}{1 + \gamma}.$$

Then for small enough  $\varepsilon$

$$\sum_{n \geq \mathcal{N}} a_n^3 n^{b+\varepsilon(3-\gamma)} < \infty. \quad (2.16)$$

Since  $f_n^{*'}(x) = -\frac{2}{\sigma^2} f_n^*(x) h_n^*(x)$  one gets from (2.14) and (2.15) that

$$\sum_{n \geq \mathcal{N}} \|T_n(F_n^*) - F_n^*\| < \infty. \quad (2.17)$$

It is easy to see that  $\|F_n^* - F_{n+1}^*\| = O(a_n^2 n^{-1})$  and, therefore

$$\sum_{n \geq \mathcal{N}} \|F_n^* - F_{n+1}^*\| < \infty. \quad (2.18)$$

Since for any distribution function  $F_n(\cdot)$



$$\|T_n(F_n) - T_n(F_n^*)\| \leq \|F_n - F_n^*\|$$

and

$$\begin{aligned} \|T_n(F_n) - F_{n+1}^*\| &\leq \|T_n(F_n) - T_n(F_n^*)\| + \\ &\|T_n(F_n^*) - F_n^*\| + \|F_n^* - F_{n+1}^*\|, \end{aligned}$$

we may take  $F_{\mathcal{N}}^*(\cdot)$  as the distribution of  $V_{\mathcal{N}}$  and sum over  $\mathcal{N} \leq k \leq n-1$  to get

$$\begin{aligned} &\|T_n(T_{n-1}(\dots T_{\mathcal{N}}(F_{\mathcal{N}}^*))) - F_n^*\| \leq \\ &\sum_{k=\mathcal{N}}^{n-1} \|T_n(F_n^*) - F_n^*\| + \sum_{k=\mathcal{N}}^{n-1} \|F_n^* - F_{n+1}^*\|. \end{aligned}$$

Due to (2.16) and (2.17), this is arbitrarily small for  $\mathcal{N}$  large enough. Hence, denoting by  $F(\cdot)$  the distribution function pertaining to the density  $f(\cdot)$ , we see that

$$\begin{aligned} &\|F - T_n(T_{n-1}(\dots T_{\mathcal{N}}(F_{\mathcal{N}}^*)))\| \leq \|F - F_n^*\| + \\ &\|F - T_n^*(T_{n-1}(\dots T_{\mathcal{N}}(F_{\mathcal{N}}^*)))\| \end{aligned}$$

is arbitrarily small. Thus the theorem is proved.

In the above theorem, the noise is given as a function of independent random variables. Another approach in the literature on stochastic approximation characterizes the noise by means of its conditional distributions. In this case, one considers a recurrent sequence

$$X_{n+1} = X_n - \gamma_n Y_n, n \geq 1, X_1 = \text{const},$$

and requires that the conditional distribution of  $Y_n$  for given  $X_1, X_2, \dots, X_n$  depends only on  $X_n$  and  $E(Y_n|X_n) = g(X_n)$ .

Set  $G(z|x) = P\{Z_n < z|X_n = x\}$ , where  $Z_n = Y_n - g(X_n)$ . We will show now that sufficient smoothness of  $G(\cdot|x)$  on  $x$  implies condition 4 of Theorem 2.1 (see also [14]).

**Corollary 2.1.** Suppose that for some  $\rho > 0$

$$\text{dist}(G(\cdot|x), G(\cdot|y)) \leq c|x - y|^\rho,$$

where  $\text{dist}(\cdot, \cdot)$  is the Levy-Prohorov distance. If  $\rho > \eta$  and

$$k = \sup_x \int |z|^\mu dG(z|x) < \infty$$

for some  $\mu > 2 + \frac{2\eta}{\rho - \eta}$ , then condition 4 holds.

**Proof.** Let  $Z_n$  be distributed according to  $G(\cdot|x)$  and  $Z_n^*$  be distributed according to  $G(\cdot|y)$ . By Strassen's well known theorem [18], there is a joint distribution for  $Z_n$  and  $Z_n^*$  such that

$$P\{|Z_n - Z_n^*| > cz^\rho\} \leq cz^\rho,$$

where  $z = |x - y|$ . Set  $\alpha = \frac{\rho-\eta}{2}$ . Consider

$$\begin{aligned} E(Z_n - Z_n^*)^2 &= E(Z_n - Z_n^*)^2 \chi_{\{|Z_n - Z_n^*| \geq cz^\rho\}} + \\ E(Z_n - Z_n^*)^2 &\chi_{\{|Z_n - Z_n^*| > cz^\rho, \max(|Z_n|, |Z_n^*|) \leq z^{-\alpha}\}} + \\ E(Z_n - Z_n^*)^2 &\chi_{\{|Z_n - Z_n^*| > cz^\rho, \max(|Z_n|, |Z_n^*|) > z^{-\alpha}\}} \leq \\ c^2 z^{2\rho} + 2cz^\rho z^{-2\alpha} + 2z^{\alpha(\mu-2)} k &\leq c_1 z^\eta. \end{aligned}$$

The corollary is proved.

**Remark 2.1** If  $X_n$  converges to 0 with probability 1, then conditions 1, 2, 4 can be replaced by their local (on  $x$ ) variants.

**Remark 2.2** If for some  $a > 0$  and  $b > 1$  the "tail"  $T(x) = 1 - F(x) + F(-x)$ ,  $x > 0$ , of an infinitely divisible distribution is  $O(\exp(-ax^b))$  as  $x \rightarrow \infty$ , then the distribution can be only a Gaussian one (including the degenerate case with zero variance) [16]. Consequently the limit in Theorem 2.1 (along with limits from [7] and [9]) is not an infinitely divisible distribution. One should not expect here convergence to an infinitely divisible distribution since the limit is generated by a nonlinear transformation of sums of small independent random variables.

Theorem 2.1 shows that the rate of asymptotic convergence increases as smoothness of the regression function (at the solution) decreases. More interesting observation can be done for the case when the regression function has several roots in which the function has different smoothness. To this end we omit the basic assumption that  $g(\cdot)$  has the unique root. Instead of this we assume that, among the roots, there are a finite number  $\Theta_i, i = 1, 2, \dots, \mathcal{N}$ , of stable ones. We call a root  $\Theta$  stable if (1.4) or (1.5) holds.

Consider the following conditions:

A. for each  $\Theta_i$  (1.4) or (1.5) holds with its own  $\alpha_j^{(i)}$ ,  $\gamma_i \in (1/2, 1]^1$ ,  $o_j^{(i)}(\cdot)$ ,  $j = 1, 2$ ;

B. if  $\gamma_i < 1$ , then

$$(a) o_j^{(i)}(x) = O(x^{1+\nu_i/\gamma_i}) \text{ for } \nu_i > \frac{1-\gamma_i}{1+\gamma_i},$$

<sup>1</sup>If (1.4) holds for  $\Theta_i$ , we set  $\gamma_i = 1$ .

(b) in a neighborhood of  $\Theta_i$

$$E|z_s(x) - z_s(\Theta_i)|^2 \leq k_i |x - \Theta_i|^{\eta_i}$$

$$\text{where } \eta_i > \frac{(1-\gamma_i)^2}{1+\gamma_i};$$

(c) for some  $\sigma_i > 0$  and  $\kappa_i > \frac{1}{2} \frac{1-\gamma_i}{1+\gamma_i}$  one has

$$\lim_{s \rightarrow \infty} |Ez_s(\Theta_i)^2 - \sigma_i^2| s^{\kappa_i} = 0;$$

(d)  $\sup_{s \geq 1} E|z_s(\Theta_i)|^{\mu_i} < \infty$  for some  $\mu_i > 2 + \frac{1-\gamma_i}{2(1+\gamma_i)(2\gamma_i-1)}$ ;

C. if  $\gamma_i = 1$ , then

(a)  $\lim_{s \rightarrow \infty} \overline{\lim}_{x \rightarrow \Theta_i} |Ez_s(x)^2 - \sigma_i^2| = 0$  for some  $\sigma_i^2 > 0$ ;

(b)  $\lim_{R \rightarrow \infty} \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{x \rightarrow \Theta_i} E|z_s(x)|^2 \chi_{\{|z_s(x)| \geq R\}} = 0$ ;

(c) either  $a\alpha_1^{(i)} = a\alpha_2^{(i)} = 1/2$  and  $o_j^{(i)}(x) = O(x^{1+\delta})$ ,  $j = 1, 2$ , for some  $\delta > 0$ , or  $2a \min(\alpha_1^{(i)}, \alpha_2^{(i)}) > 1$ .

**Theorem 2.2.** Suppose that the sequence  $\{X_n\}$  given by (1.3) converges with probability 1 and conditions A, B, C hold. Then

$$\lim_{n \rightarrow \infty} P\{r_n^{(i)}(X_n - \Theta_i) < x, \lim_{s \rightarrow \infty} X_s = \Theta_i\} = \mathcal{F}_i(x) P\{\lim_{s \rightarrow \infty} X_s = \Theta_i\}.$$

Here

$$r_n^{(i)} = \begin{cases} n^{1/1+\gamma_i} & \text{if } \gamma_i < 1, \\ \sqrt{n} & \text{if } \gamma_i = 1 \text{ and } 2a \min(\alpha_1^{(i)}, \alpha_2^{(i)}) > 1, \\ \sqrt{\frac{n}{\ln n}} & \text{if } \gamma_i = 1 \text{ and } a\alpha_1^{(i)} = a\alpha_2^{(i)} = 1/2. \end{cases}$$

Also  $\mathcal{F}_i(\cdot)$  stands for a distribution function such that:

a) for  $\gamma_i < 1$

$$\mathcal{F}_i'(x) = c_i \begin{cases} \exp\left\{-\frac{2\alpha_1^{(i)}}{a\sigma_i^2} \frac{x^{1+\gamma_i}}{1+\gamma_i}\right\}, & x \geq 0, \\ \exp\left\{-\frac{2\alpha_2^{(i)}}{a\sigma_i^2} \frac{|x|^{1+\gamma_i}}{1+\gamma_i}\right\}, & x < 0, \end{cases}$$

b) for  $\gamma_i = 1$  and  $2a \min(\alpha_1^{(i)}, \alpha_2^{(i)}) > 1$

$$\mathcal{F}_i'(x) = c_i \begin{cases} \exp\left\{-\frac{2a\alpha_1^{(i)}-1}{2a^2\sigma_i^2} x^2\right\}, & x \geq 0, \\ \exp\left\{-\frac{2a\alpha_2^{(i)}-1}{2a^2\sigma_i^2} x^2\right\}, & x < 0; \end{cases}$$

c) for  $\gamma_i = 1$  and  $a\alpha_1^{(i)} = a\alpha_2^{(i)} = 1/2$

$$\mathcal{F}_i'(x) = \frac{1}{\sqrt{2\pi a^2 \sigma_i^2}} \exp\left\{-\frac{x^2}{2a^2 \sigma_i^2}\right\},$$

where  $c_i$  stands for a normalizing constant.

**Proof.** Set

$$\tilde{g}(x) = \begin{cases} g(x) & \text{for } |x - \Theta_i| \leq \varepsilon_i, \\ \alpha_i(x - \Theta_i) & \text{for } |x - \Theta_i| > \varepsilon_i; \end{cases}$$

$$\tilde{z}_s(x) = \begin{cases} z_s(x) & \text{for } |x - \Theta_i| \leq \varepsilon_i, \\ z_s(\Theta_i) & \text{for } |x - \Theta_i| > \varepsilon_i; \end{cases}$$

$$\tilde{X}_{n+1}^{N,y} = \tilde{X}_n^{N,y} - \gamma_n[\tilde{g}(\tilde{X}_n^{N,y}) + \tilde{z}_n(\tilde{X}_n^{N,y})], n \geq N, \tilde{X}_N^{N,y} = y.$$

Here  $\alpha_i > 0$  and  $\varepsilon_i$  is so small that the condition b) from B holds. Also  $y$  does not depend on  $z_s(x), s \geq N$ . By Theorem 2.1 or corresponding results from [4, 7, 9, 10, 11, 17]

$$\lim_{n \rightarrow \infty} P\{r_n^{(i)}(\tilde{X}_n^{N,y} - \Theta_i) < x\} = \mathcal{F}_i(x). \quad (2.19)$$

Introduce the events  $A_{n,\delta} = \{|X_n - \Theta_i| < \delta\}$  and  $B_{n,\delta} = \{|X_s - \Theta_i| < \delta, s \geq n\}$ , where  $n \geq 1, \delta \in (0, 1)$ . By hypothesis  $X_n$  converges with probability 1. Therefore, for any  $\sigma > 0$  we find  $\delta$  and  $n(\delta)$  such that for  $n \geq n(\delta)$

$$P\{\{\lim_{s \rightarrow \infty} X_s = \Theta_i\} \Delta B_{n,\delta}\} < \sigma$$

and

$$P\{A_{n,\delta} \Delta B_{n,\delta}\} < \sigma.$$

Here the sign  $\Delta$  denotes the symmetric difference.

Using (2.18), the Markovian property and the Lebesgue Dominated Convergence Theorem, we have for  $n \geq n(\delta)$

$$\begin{aligned} & \overline{\lim}_{m \rightarrow \infty} P\{r_m^{(i)}(X_m - \Theta_i) < x, \lim_{s \rightarrow \infty} X_s = \Theta_i\} \leq \\ & \overline{\lim}_{m \rightarrow \infty} P\{r_m^{(i)}(X_m - \Theta_i) < x, B_{n,\delta}\} + \sigma = \\ & \overline{\lim}_{m \rightarrow \infty} P\{r_m^{(i)}(\tilde{X}_m^{n,X_n} - \Theta_i) < x, B_{n,\delta}\} + \sigma \leq \\ & \overline{\lim}_{n \rightarrow \infty} P\{r_m^{(i)}(\tilde{X}_m^{n,X_n} - \Theta_i) < x, A_{n,\delta}\} + \sigma = \\ & \overline{\lim}_{n \rightarrow \infty} EP\{r_m^{(i)}(\tilde{X}_m^{n,X_n} - \Theta_i) < x | X_n\} \chi_{A_{n,\delta}} + \sigma = \\ & \mathcal{F}_i(x) P\{A_{n,\delta}\} + \sigma \leq \mathcal{F}_i(x) P\{\lim_{s \rightarrow \infty} X_s = \Theta_i\} + 3\sigma. \end{aligned}$$

Similarly,

$$\underline{\lim}_{n \rightarrow \infty} P\{r_m^{(i)}(X_m - \Theta_i) < x, \lim_{s \rightarrow \infty} X_s = \Theta_i\} \geq$$

$$\mathcal{F}_i(x)P\left\{\lim_{s \rightarrow \infty} X_s = \Theta_i\right\} - 3\sigma.$$

Since  $\sigma$  is arbitrarily small, these inequalities yield the required result. The theorem is proved.

**Remark 2.2.** Conditions which ensure positivity of  $P\{\lim_{n \rightarrow \infty} X_n = \Theta_i\}$  are known [8].

Suppose now that the process (1.3) converges with positive probability to each of the stable roots (which are called attainable) and to all other roots (which are called in this case unattainable) with zero probability (see for particular cases of this (touchpoints and unstable points) [11] and [13] correspondingly). Then

$$\sum_{i=1}^{\mathcal{N}} P\{X_n \rightarrow \Theta_i\} = 1$$

and the asymptotic behavior of our process can be imagined in the following way. By chance one selects a stable point (to which the process will converge) and a “convergence mechanism” (depending upon the local properties of the process at the point) switches on to drive the process to the point.

Theorem 2.2 covers only the cases when limit distributions are “nice” (remember that, except Gaussian, the distributions are not infinitely divisible). Other cases known in stochastic approximation [10, 11] can be treated in the same way.

### 3 COMPLEX LIMIT BEHAVIOR IN URN SCHEMES AND TECHNOLOGICAL DYNAMICS

Returning to the model of technological dynamics presented in section 1, we can find, using the given results, the rates of convergence to the two stable roots.

We give now a lemma which ensures reformulation of the above theorems for the generalized urn scheme.

Let  $\tau_i, i \geq 1$ , be independent random variables uniformly distributed on  $[0, 1]$ . Set

$$\zeta_i(x) = \chi_{\{\tau_i < x\}}, i \geq 1, x \in [0, 1].$$

Elementary manipulations ensure the following result.

**Lemma 3.1** One has  $E[\zeta_i(x) - \zeta_i(y)]^2 = x + y - 2 \min(x, y) \leq |x - y|$ .

Also

$$\zeta_i(x) = \begin{cases} 1 & \text{with probability } x, \\ 0 & \text{with probability } 1 - x. \end{cases}$$

Now designating  $\zeta_i(f(x))$  by  $\xi_i(x)$  and using the recursion (1.3), we can derive analogs of the above theorems for the generalized urn scheme.

Returning to the urn scheme formulations, for the model of technological dynamics there are three roots  $\theta_1, \theta_2$  and  $\theta_3$  of the function  $f(x) - x$  on  $[0, 1]$ . The root  $\theta_2$  proves to be unattainable, while  $\theta_1$  and  $\theta_3$ , being stable, are attainable, i.e. the process  $X_t$  converges to each of them with positive probability for any initial amounts of the technologies on the market. Using the above results we can find the rates of convergence to the both attainable roots

$$\theta_1 = \frac{(\alpha + \Delta_1)(1 - x_B^*)}{2\alpha(1 - x_B^*) + \Delta_1 + \Delta_2},$$

$$\theta_3 = 1 - \frac{(\alpha + \Delta_4)(1 - x_A^*)}{2\alpha(1 - x_A^*) + \Delta_3 + \Delta_4}.$$

In particular, for every  $y$

$$\lim_{t \rightarrow \infty} P\{\sqrt{t}(X_t - \theta_i) < y, X_s \rightarrow \theta_i\} = P\{X_s \rightarrow \theta_i\}P\{\mathcal{N}(0, \sigma_i^2) < y\}.$$
 (3.1)

Here  $\mathcal{N}(0, \sigma_i^2)$  stands for a Gaussian distribution with zero mean and variance

$$\sigma_i^2 = \frac{\theta_i(1 - \theta_i)}{1 - 2f'(\theta_i)},$$
 (3.2)

where  $f'(\cdot)$  denotes the derivative of  $f(\cdot)$ . It can be shown that

$$f'(\theta_1) = -\frac{\Delta_1 + \Delta_2}{2\alpha(1 - x_B^*)}$$
 (3.3)

and

$$f'(\theta_3) = -\frac{\Delta_3 + \Delta_4}{2\alpha(1 - x_A^*)}.$$
 (3.4)

From (3.1) – (3.4) one sees that convergence to both  $\theta_1$  and  $\theta_3$  occurs with the rate  $t^{-1/2}$  but the random fluctuations (which are determined by the variances of the corresponding limit distributions) around this dominant tendency can be different.

Now we conclude that the above dynamics of prices, together with the described behavior of adopters, generate multiple limit patterns with slightly different rates of convergence. With such dynamics of prices and assumptions concerning the behavior of adopters, one can have even more

complicated limit market structure where the ratio between the initial amounts of the technologies on the market influences the structure [3]. Also the considerations concerning the convergence rate are applicable (with corresponding modifications) to the more complicated situation analysed in [3].

The situation with deterministic choice can also be treated within this framework. It corresponds to  $\xi \equiv 0$ . In this case for any initial proportion belonging to  $(0, \gamma_2]$  one has (deterministic) convergence to  $\gamma_1$  with the rate of the order of  $t^{-1}$  (more precisely  $\limsup_{t \rightarrow \infty} t|X_t - \gamma_1| = \max(\gamma_1, 1 - \gamma_1)$ ). Also for initial proportions belonging to  $(\gamma_2, 1)$  – convergence to  $\gamma_3$  with the same rate. Consequently in this case the limit behavior is essentially the same as previously, but the limit depends upon the initial state and the convergence is faster.

Formally speaking the main difference between the deterministic choice and the random choice is that the corresponding measures (search for alternatives) are concentrated at one point or in some non-singleton set (usually an interval). The above measure corresponding to the uniform distribution, is proportional to the Lebesgue measure. The intermediate case between the measure concentrated at one point and a measure uniformly distributed among an interval, would be a measure such that for some interval  $[a, b]$

$$\frac{\text{mes}([a, b])}{\text{mes}_L([a, b])} \rightarrow \infty \quad (3.5)$$

as  $a \rightarrow b$ . Here  $\text{mes}(\cdot)$  stands for the measure under consideration and  $\text{mes}_L(\cdot)$  – for the Lebesgue measure on the real line. As an example consider the following distribution function for  $\xi$

$$G(x) = \begin{cases} 1/2 + a_1 x^\gamma, & 0 \leq x \leq (1/2a_1)^{1/\gamma}, \\ 1/2 - a_2 |x|^\gamma, & -(1/2a_2)^{1/\gamma} \leq x < 0, \end{cases} \quad (3.6)$$

where  $a_1 > 0$ ,  $a_2 > 0$ ,  $\gamma < 1$ . Then for any  $[a, b]$  containing 0 property (3.5) holds. The above decision rule with this particular distribution function of  $\xi$  does not bring in general anything new into our considerations. We still have the convergence rate of order  $t^{-1/2}$  as  $t \rightarrow \infty$  to both attainable roots  $\theta_1$  and  $\theta_3$ , with Gaussian distributions in the corresponding formulae similar to (3.1). But if one of the attainable roots  $\theta_1$  or  $\theta_3$  coincides with  $1/2$ , we can observe a substantially different convergence rate to it. This can be the case when the dynamics of prices has the form given by Figure 3. Conceptually this situation means that the producers of  $B$  are so “optimistic” that switch from lowering to increasing the price of  $B$  until the level of market share equals to  $1/2$ .

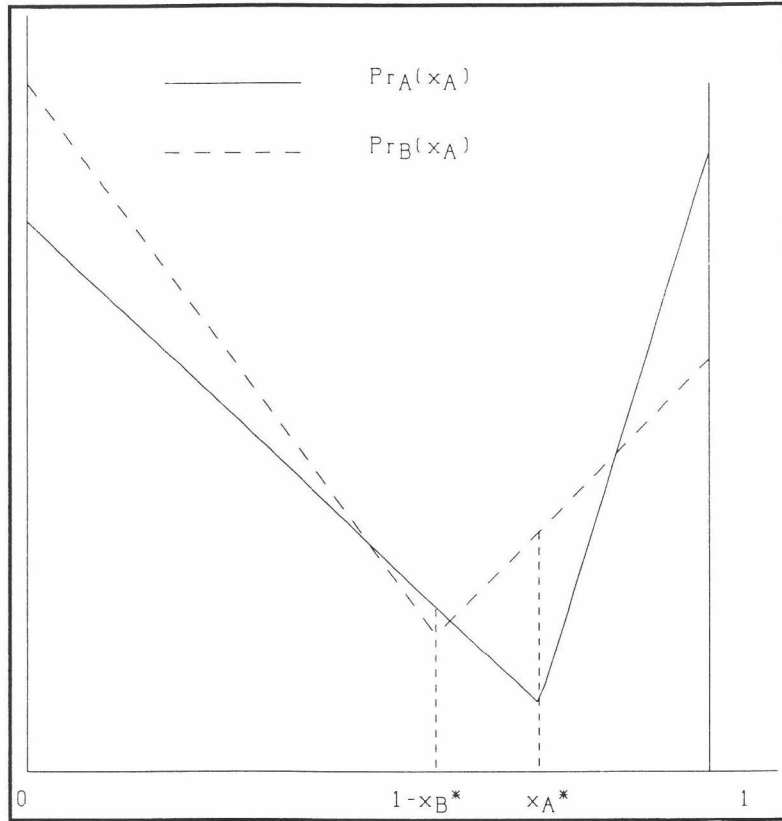


FIG. 3. Dynamics of prices of  $A$  and  $B$  in the case of optimistic expectations of the producers of  $B$ .

The corresponding function  $f(\cdot)$  is given on Figure 4. Here  $\theta_1 = 1/2$  and  $\theta_3$  are attainable roots and  $\theta_2$  is unattainable. Using the results of section 2 we have for every  $y$

$$\begin{aligned} \lim_{t \rightarrow \infty} P\{\sqrt{t}(X_t - \theta_3) < y, X_s \rightarrow \theta_3\} = \\ P\{X_s \rightarrow \theta_3\}P\{\mathcal{N}(0, \sigma_3^2) < y\} \end{aligned} \quad (3.7)$$

and

$$\lim_{t \rightarrow \infty} P\{t^{1/1+\gamma}(X_t - 1/2) < y, X_s \rightarrow 1/2\} = P\{X_s \rightarrow 1/2\}F(y), \quad (3.8)$$

where

$$\sigma_3^2 = \frac{\theta_3(1 - \theta_3)}{1 - 2f'(\theta_3)},$$



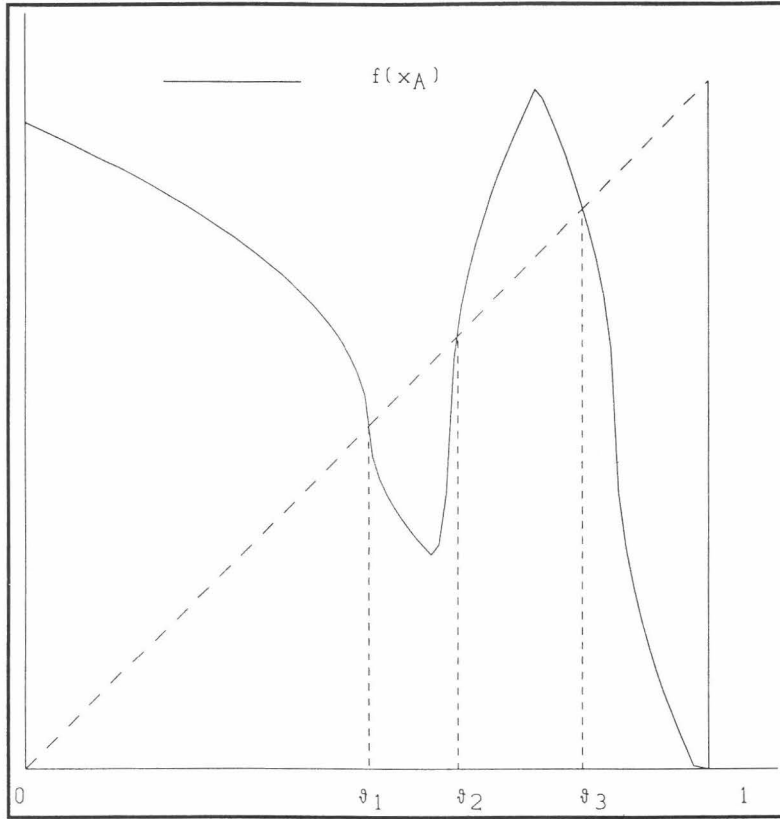


FIG. 4. Probability of choosing  $A$  depending on its market share in the case when the “search-factor”  $\xi$  has the distribution of form (3.6).

$$f'(\theta_3) = -\gamma a_1 [Pr_A(\theta_3) - Pr_B(\theta_3)]^{\gamma-1} \frac{\Delta_3 + \Delta_4}{(1 - x_A^*)}$$

$$F'(x) = C \begin{cases} \exp \left\{ -\frac{2^{5+\gamma} a_1 \Delta_1^\gamma}{1+\gamma} x^{1+\gamma} \right\}, & x \geq 0, \\ \exp \left\{ -\frac{2^{5+\gamma} a_2 \Delta_1^\gamma}{1+\gamma} (-x)^{1+\gamma} \right\}, & x < 0, \end{cases}$$

and  $C$  is a normalizing constant. Also relation (3.8) follows from the results of Section 2 only for  $\gamma > 1/2$ . In contrast to the previous case, we have faster rate of convergence,  $t^{1/\gamma+1}$  (to the root  $1/2$ ). This corresponds to the intermediate situation between deterministic and regular random search for alternatives. What is very important conceptually, is that for the same process we can observe different convergence rates for different attainable limit states. The limit market structure corresponding to the

root  $\theta_1 = 1/2$  will be faster emerging than the structure corresponding to  $\theta_3$ , although both are feasible.

If we introduce certain hypothesis about statistical frequencies (probabilities) of the producers of  $A(B)$  who follow a particular shape of the above price dynamics and/or hypothesis concerning statistical frequencies of the adopters who use a concrete variant of the above decision making mechanism, we can construct much more complicated functions  $f(\cdot)$ .

#### 4 CONCLUSIONS

Not only the diffusion of innovations studies [1, 3], but also other important problems such as autocatalytic chemical reactions [1, 12] and some biological populations [6] can be treated within the framework of the generalized urn scheme. In these conceptual problems the limit theorems given before serve as a means for the analysis of the rates of convergence to attainable components of the terminal set (which resemble rates of formation of the final market shares in the diffusion of innovation studies or the rates of conversion of initial ingredients into the final products in the autocatalytic chemical reactions or the rates of origination of new species in the biological studies). The results show that in the case with multiple limit states<sup>2</sup>, the rates are different and depend upon the smoothness of the urn function  $f(\cdot)$  in neighborhood of the states. The rate of development of the predominant trend, in general, does not exist for a process with multiple limit states – some of the tendencies develop quicker, others slower.

All phenomena mentioned above demonstrate the essential nonlinearity of the stochastic processes generated by the generalized urn scheme in the case of multiple equilibria. Also one can see that the theorems given here represent a powerful and convenient tool for studying and demonstrating the nonlinear effects pertinent to the processes.

**Acknowledgements.** The authors are grateful to the two anonymous referees and the editor for their careful review of this paper.

#### REFERENCES

- [1] Arthur, W.B., Yu. M. Ermoliev and Yu. M. Kaniovski (1987). Path Dependent Processes and the Emergence of Macro-Structure. *European Journal of Operational Research*, **30**, pp. 294–303.
- [2] Burkholder, D.L. (1956). On a Class of Stochastic Approximation Procedures. *Ann. Math. Stat.*, **25**, pp. 1044–1059.

---

<sup>2</sup>We did not consider here the case when there are intervals among the limit states (see [5].)

- [3] Dosi, G., Yu. Ermoliev and Yu. Kaniovski (1994). Generalized Urn Schemes and Technological Dynamics. *Journal of Mathematical Economics*, **23**, pp. 1–19.
- [4] Fabian, V. (1968). On Asymptotic Normality in Stochastic Approximation. *Ann. Math. Stat.*, **39**, pp. 1327–1332.
- [5] Hill, B.M., D. Lane and W. Sudderth (1980). A Strong Law for Some Generalized Urn Processes. *Ann. Prob.*, **8**, pp. 214–226.
- [6] Hofbauer, F., and Sigmund, K. (1988). *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection*, Cambridge University Press, Cambridge.
- [7] Kaniovskaia, I.Yu. (1979). Limit Theorems for Recurrent Adaptation Algorithms with Non-Smooth Regression Functions. *Probabilistic Methods in Cybernetics*, Kiev, pp. 57–65 (Preprint of Institute of Cybernetics of the Ukrainian SSR Academy of Sciences No. 79–69) (in Russian).
- [8] Kaniovski, Yu. M. (1988). Limit Theorems for Processes of Stochastic Approximation when the Regression Function has Several Roots. *Kibernetika*, No. 2, pp. 136–138 (in Russian).
- [9] Kersting, G.D. (1978). A Weak Convergence Theorem with Application to the Robbins-Monro Process. *Ann. Prob.*, **6**, pp. 1015–1025.
- [10] Major, P., and P. Revesz (1973). A Limit Theorem for the Robbins-Monro Approximation. *Z. Wahrsch. Verw. Geb.*, **27**, pp. 79–86.
- [11] Nevel'son, M. and R. Has'minski (1972). *Stochastic Approximation and Recurrent Estimation*, Nauka, Moscow (in Russian).
- [12] Nicolis, G., and I. Priogogine (1971). *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*, Wiley, New York.
- [13] Pemantle, R. (1991). When Are Touchpoints Limits for Generalized Polya Urns. *Proceedings of the American Mathematical Society*, **113**, pp. 235–243.
- [14] Pflug, G. (1985). The Robbins – Monro Procedure in Nonstandard Situations. Unpublished manuscript of the University of Giessen.
- [15] Robbins, H. and S. Monro (1951). A Stochastic Approximation Method. *Ann. Math. Statist.*, **22**, pp. 400–407.

- [16] Ruegg, A. (1970). A Characterization of Certain Infinitely Divisible Laws. *Ann. Math. Stat.*, **41**, pp. 1354–1356.
- [17] Sacks, J. (1958). Asymptotic Distribution of Stochastic Approximation Procedures. *Ann. Math. Stat.*, **29**, pp. 375–405.
- [18] Strassen, V. (1965). The Existence of Probability Measures with Given Martingals. *Ann. Math. Statist.*, **36**, pp. 423–439.

Received: 2/15/1993  
Revised: 4/13/1994  
Accepted: accepted 6/30/1994

Recommended by Simon Tavaré, Editor