Working Paper

Non-standard Limit Theorems for Stochastic Approximation Procedures and Their Applications for Urn Schemes

Yu. Kaniovski
G. Pflug

WP-92-25
March 1992

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Non-standard Limit Theorems for Stochastic Approximation Procedures and Their Applications for Urn Schemes

Yu. Kaniovski
G. Pflug

WP-92-25
March 1992

IIASA International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 715210 ◻ Telex: 079 137 iiasa a ◻ Telefax: +43 2236 71313

Foreword

A limit theorem for the Robbins-Monro stochastic approximation procedure is proved in the case of a non-smooth regression function. Using this result a conditional limit theorem is given for the case when the regression function has several stable roots. The first result shows that the rate of convergence for the stochastic approximation-type procedures (including Monte-Carlo optimization algorithms and adaptive processes of growth being modelled by the generalized urn scheme) decreases as the smoothness increases. The second result demonstrates that in the case of several stable roots, there is no convergence rate for the procedure as whole, but for each of stable roots there exists its specific rate of convergence. The latter allows to derive several conceptual results for applied problems in biology, physical chemistry and economics which can be described by the generalized urn scheme.

Contents

Non-standard Limit Theorems for Stochastic Approximation Procedures and Their Applications for Urn Schemes

*Yu. Kaniovski**

*G. Pflug***

1 Introduction

Consider the Robbins-Monro procedure [1] for finding the root of a (Borel measurable) regression function $g(x)$, $x \in \mathbf{R}$, $\mathbf{R}$ - the set of real numbers. Supose we have

$$y(s,x) = g(x) + z_s(x), \quad s \geq 1, \quad x \in \mathbf{R}, \tag{1.1}$$

independent (in s) observations of $g(.)$ with a random noise $z_s(\cdot)$. Here, $z_s(x)$ stands for a random field (on $\mathbf{N} \times \mathbf{R}, \mathbf{N}$ — the set of natural numbers and on some fixed probability space $(\Omega, \mathcal{F}, P)$) with zero mean, i.e. $\mathrm{E}z_s(x) = 0$; finite variances, i.e. $\mathrm{E}z_s(x)^2 = \sigma_s^2(x)$; independent in the first argument values, i.e. $z_s(x)$ and $z_n(y)$ are independent for $s \neq n$ for any (deterministic) $x, y \in \mathbf{R}$. Also, $z_n(.)$ is a measurable mapping with respect to $\mathcal{B} \times \mathcal{F}$ (by $\mathcal{B}$ we designate the σ-algebra of Borel sets in $\mathbf{R}$). Then the Robbins-Monro procedure gives successive approximations $X_n, n \geq 1$, to the root Θ in the following form:

$$X_{n+1} = X_n - \gamma_n y(n, X_n), \quad X_1 = \mathrm{const}, \tag{1.2}$$

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

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

*IIASA, Laxenburg, Austria

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

For the case when $\gamma_n = an^{-1}$ asymptotic normality of $\sqrt{n}(X_n - \Theta)$ was proved [2, 4, 11, 15] for locally linear $g(\cdot)$ at Θ, i.e. $g(x) = g'(\Theta)(x - \Theta) + o(|\, x - \Theta\,|)$ as $x \to \Theta$, and $2\, ag'(\Theta) > 1$. Also $\sqrt{\frac{n}{\ln n}}(X_n - \Theta)$ is asymptotically normal if $2\, ag'(\Theta) = 1$ [10]. But for the simplest case, when $g(.)$ is not locally linear at Θ, i.e., as $x \to \Theta$

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

$\sqrt{n}(X_n - \Theta)$ converges weakly for $2a \min(\alpha_1, \alpha_2) > 1$ to a non-Gaussian limit distribution [7,9].

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

$$g(x) = \begin{cases} \alpha_1(x - \Theta)^\gamma \ + \ o_1((x - \Theta)^\gamma), & x \geq \Theta, \\ -\alpha_2(\Theta - x)^\gamma \ + \ o_2((\Theta - x)^\gamma), & x < \Theta, \end{cases} \tag{1.3}$$

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 (1.2) and in the generalized urn scheme demonstrate complex limit behaviour caused by both the nonlinearity of the form (1.3) of functions involved and the multiplicity of their roots. Under the 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 $t = 1, 2, \ldots$. 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. The function $f(.)$ is called urn function for this generalized urn scheme (see, for example, [5]). Designate by X_t the proportion of black balls into the urn at time t. Let for $t \geq 1$ and $x \epsilon \mathrm{R}\ [0, 1]$

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

be independent in t. Here R[0,1] stands for the set of rational numbers from [0, 1]. Then $\{X_t\}$ follows the dynamics

$$X_{t+1} = X_t + \frac{1}{\gamma_1 + t}[\xi_t(X_t) - X_t] = \quad (1.4)$$

$$X_t + \frac{1}{\gamma_1 + t}[f(X_t) - X_t] + \quad (1.5)$$
$$\frac{1}{\gamma_1 + t}\eta_t(X_t), \quad t \geq 1, \quad X_1 = b_1/\gamma_1,$$

where $\eta_t(x) = \xi_t(x) - f(x)$. Taking into account (1.1), one can see that (1.4) represents a recurrent relation of the form (1.2) with $g(x) = f(x) - x$ and $z_t(x) = \eta_t(x)$. Consequently, both (1.2) and (1.4) can be studied with the same machinery.

The generalized urn scheme proves to be a convenient tool for modelling of complex phenomena in economics and biology [1, 3].

Now we proceed to limit theorems for random variables generated by (1.2).

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

We study the algorithm (1.2) with $\Theta = 0$ and $\gamma_n = an^{-1}$, $a > 0$. For a real valued function $h(.)$ 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. $g(.)$ has the form (1.3) and $|o_i(y)| = O(y^\nu)$ 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 $\kappa > \frac{1}{2}\frac{1-\gamma}{1+\gamma}$ one has $\lim_{s\to\infty} |\mathrm{E}z_s(0)^2 - \sigma^2|s^\kappa = 0)$;

6. $\sup_s \mathrm{E}|z_s(0)|^\mu < \infty$ for some $\mu > 2 + \frac{1-\gamma}{2(1+\gamma)(2\gamma-1)}$.

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

$$n^{\frac{1}{1+\gamma}} X_n \underset{n\to\infty}{\overset{w}{\longrightarrow}} X,$$

where X has the density

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

Here C stands for a normalizing constant.

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

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

where

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

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

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

We shall construct functions $h_n^*(u)$ with the following properties

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

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

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

$$a_n^2 \inf_{|u|\le k_n} |h_n^{*'}(u)| \ge cn^{-1/2\beta}, \tag{2.5}$$

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

where ε will be fixed later. To this end, let $h_n^*(.)$ be a smoothed modification of

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

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

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

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

$$\inf_{\|u\|\le k_n} h_n^{*'}(u) \ge c_1 n^{-\frac{\beta}{2}(1-\gamma)^2}.$$

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

$$|n^{1-\beta} g(n^{-\beta}u) - \alpha_1 |u^\gamma|| = n^{1-\beta} O((n^{-\beta}u)^{\gamma+\nu})$$

$$= \quad O(n^{1-\beta+(\gamma+\nu)(-\frac{1}{2}+\delta)}) = O(n^{-\xi}).$$

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

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

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

$$W_{n+1} = W_n - a_n^2 h_n^*(W_n) + a_n z_n, n \geq \mathcal{N}, \tag{2.7}$$

with $W_{\mathcal{N}}$ — arbitrary (but it does not depend on $z_n(x), n \geq 1$, for any (deterministic) x).

Let τ 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 clouse to 1 by choosing $\mathcal{N}$ large. On the event $\{\tau = \infty\}$, using the bound (2.5), we get

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

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

$$V_{n+1} = V_n - a_n^2 h_n^*(V_n) + a_n z_n^{'}, n \geq \mathcal{N}, V_{\mathcal{N}} - \text{arbitrary}, \tag{2.8}$$

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

$$\mathrm{E}[(z_n - z_n^{'})^2 \chi_{\{\tau' > n\}}] \leq k\mathrm{E}[(n^{\beta}|W_n|)^{\eta} . \chi_{\{\tau' > n\}}] \leq kn^{(-\frac{1}{2}+\delta)\eta}. \tag{2.9}$$

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

$$|h_n^*(u) - h_n^*(v)| \leq c_2 + c_3|u - v|. \tag{2.10}$$

If δ is so small that $(\frac{1}{2} - \delta)\eta > \frac{\beta}{2}(1-\gamma)^2$, then from (2.5), (2.9) and (2.10) we have

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

Hence due to Lemma 2.1

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

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

Due to condition 5, the arguments identical to the ones given above show that the recursion of the form (2.8) with z_n' substituted by $\sigma[\mathrm{E}z_n(0)^2]^{-1/2}z_n(0)$ has the same limit distribution. Consequently, we can suppose that $\mathrm{E}(z_n')^2 = \sigma^2, n \geq \mathcal{N}$.

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

$$b > \frac{1-\gamma}{2(1+\gamma)^2(\mu - 2)}. \tag{2.11}$$

By Markov's inequality

$$|Ez_n''| \leq n^{b(1-\mu)}E|z_n'|^\mu \tag{2.12}$$

and

$$|\mathrm{E}z_n''|^2 \le n^{b(2-\mu)}\mathrm{E}|z_n'|^{\mu}.$$

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

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

$$|\mathrm{E}(z_n'')^2 - \sigma^2| \le n^{b(2-\mu)}\mathrm{E}|z_n'|^{\mu}$$

and

$$\mathrm{Var}\ z_n''' = \mathrm{Var}\ z_n'' = \mathrm{E}(z_n'')^2 - (\mathrm{E}Z_n'')^2$$

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

From now on we consider the recursion

$$\mathrm{V}_{n+1} = \mathrm{V}_n - a_n^2 h_n^*(\mathrm{V}_n) - a_n z_n^*, n \ge \mathcal{N},$$

where $\mathrm{V}_{\mathcal{N}}$ -arbitrary (but it does not depend on $z_n^*, n \ge \mathcal{N}$). Notice that for large enough n

$$|z_n^*| \le 2n^b \text{ a.s.}, \mathrm{E}z_n^* = 0,\ \mathrm{Var}\ z_n^* = \sigma^2.$$

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

$$|H_n^{-1}(x) - [x + a_n^2 h_n^*(x)]\,| \le 2|x - H_n(x + a_n^2 h_n^*(x))\,| \le \tag{2.13}$$

$$\le a_n^4 \|h_n^*\|\ \|h_n^{*'}\| = O(a_n^3).$$

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

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

where $G_n(.)$ is the distribution function of z_n^*. Let $F_n^*(.)$ be the distribution with density

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

where C_n is a normalizing constant. We show that $T_n(F_n^*)(.)$ is close to $F_n^*(.)$, i.e. $F_n^*(.)$ 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)

$$\begin{aligned} T_n(F_n^*)(x) &= \int F_n^*(z + a_n^2 h_n^*(z))dG_n(\frac{x-z}{a_n}) + O(a_n^3) \\ &= F_n^*(x + a_n^2 h_n^*(x)) + \textstyle\int (x-z)\frac{\partial}{\partial x}[F_n^*(x + a_n^2 h_n^*(x))]dG_n(\frac{x-z}{a_n}) \\ &+ \tfrac{1}{2}\textstyle\int (x-z)^2 \frac{\partial^2}{\partial x^2}[F_n^*(x + a_n^2 h_n^*(x))]dG_n(\frac{x-z}{a_n}) \\ &+ \tfrac{1}{6}\textstyle\int (x-z)^3 \frac{\partial^3}{\partial x^3}[F_n^*(\tilde{x} + a_n^2 h_n^*(\tilde{x}))]dG_n(\frac{x-z}{a_n}) + O(a_n^3) \\ &= F_n^*(x + a_n^2 h_n^*(x)) + \tfrac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2}[F_n^*(x + a_n^2 h_n^*(x))] \\ &+ O(a_n^3 n^b n^{\varepsilon(3-\gamma)}) + O(a_n^3) = F_n^*(x) + a_n^2 f_n^*(x) h_n^*(x) \\ &+ a_n^2 \tfrac{\sigma^2}{2} f_n^{*\prime}(x) + O(a_n^3 n^{b+\varepsilon(3-\gamma)}). \end{aligned} \tag{2.14}$$

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

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

Then for small enough ε

$$\sum_{n \geq \mathcal{N}} a_n^3 n^{b+\varepsilon(3-\gamma)} < \infty. \tag{2.15}$$

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

$$\sum_{n \geq \mathcal{N}} \|T_n(F_n^*) - F_n^*\| < \infty. \tag{2.16}$$

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

$$\sum_{n \geq \mathcal{N}} \|F_n^* - F_{n+1}^*\| < \infty. \tag{2.17}$$

Since for any distribution function $F_n(.)$

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

and

$$\|T_n(F_n) - F^*_{n+1}\| \le \|T_n(F_n) - T_n(F^*_n)\|$$
$$+ \quad \|T_n(F^*_n) - F^*_n\| + \|F^*_n - F^*_{n+1}\|,$$

we may take $F^*_{\mathcal{N}}(.)$ as the distribution of $V_{\mathcal{N}}$ and sum over $\mathcal{N} \le k \le n-1$ to get

$$\|T_n(T_{n-1}(\ldots T_{\mathcal{N}}(F^*_{\mathcal{N}}))) - F^*_n\| \le \sum_{k=\mathcal{N}}^{n-1} \|T_n(F^*_n) - F^*_n\| + \sum_{k=\mathcal{N}}^{n-1} \|F^*_n - F^*_{n+1}\|.$$

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

$$\|F - T_n(T_{n-1}(\cdots T_{\mathcal{N}}(F^*_{\mathcal{N}})))\| \le \|F - F^*_n\| + \|F - T^*_n(T_{n-1}(\cdots T_{\mathcal{N}}(F^*_{\mathcal{N}})))\|$$

is arbitrary small. Thus the theorem is proved.

In the above theorem, the noise is given as a random field with certain statistical structure. 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

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

and requires that the conditional distribution of Y_n for given $X_1, X_2, \ldots, X_n$ 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 $G(.|x)$ on x implies condition 4 of theorem 2.1.

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

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

where dist (.,.) 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 $G(.|x)$ and Z_n' be distributed according to $G(.|y)$. By Strassen's well known theorem [16], 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

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

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.

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 ommit the basic for stochastic approximation assumption that $g(.)$ 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 here a root Θ stable if (1.3) holds.

Consider the following conditions:

A. for each Θ_i (1.3) holds with its own $\alpha_j^{(i)}$, $\gamma_i \epsilon (1/2, 1]$, $o_j^{(i)}(.)$, $j = 1, 2$;

B. if $\gamma_i < 1$, then

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

(b) in a neighbourhood of Θ_i

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

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

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

C. if $\gamma_i = 1$, then

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

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

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

or $2 \min (\alpha_1^{(i)}, \alpha_2^{(i)}) > 1$

Theorem 2.2. Suppose that the sequence $\{X_n\}$ given by (1.2) converges with probability 1 and conditions A, B, C hold. Then

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

Here

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

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

a) for $\gamma_i < 1$

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

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

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

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

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

where c_i stands for a normalizing constant.

Proof. Set

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

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

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

Here $\alpha_i > 0$ and ε_i is so small that the condition b) from B holds. Also y does not depend on $z_s(x), s \ge N$, for any (deterministic) x. By theorem 2.1 or corresponding results from [4, 7, 9, 10, 11, 15]

$$\lim_{n\to\infty} P\{r_n^{(i)}(\tilde{X}_n^{N,y} - \Theta_i) < x\} = \mathcal{F}_i(x). \tag{2.18}$$

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

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

and

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

Here the sign Δ denotes the symmetric difference.

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

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

Similarly,

$$\lim_{n\to\infty} P\left\{r_m^{(i)}(X_m - \Theta_i) < x, \lim_{s\to\infty} X_s = \Theta_i\right\}$$
$$\geq \mathcal{F}_i(x) P\left\{\lim_{s\to\infty} X_s = \Theta_i\right\} - 3\sigma.$$

Since σ is arbitrary small, these inequalities yield the required result.

The theorem is proved.

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

Suppose now that the process (1.2) converges with positive probability to each of stable roots and to all other roots with zero probability (see for particular cases of this (touchpoints and unstable points) [11] and [13] correspondingly). Then

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

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 "plausable" (note that, except Gaussian, the distributions are not infinitely divisible). Other cases known in stochastic approximation [10, 11] can be treated in the same way.

More interesting conceptual examples come from applications of the generalized urn scheme.

3 Limit Distributions for the Generalized Urn Scheme in Non-standard Cases

Some practically important problems in the diffusion of innovations studies [1, 3], in the autocatalytic chemical reactions [1, 12] and in the analysis of dynamics of 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 the rates of formation of the final market shares in the diffusion of innovation studies or the rates of convertion of initial ingredients into the final products in the autocatalitic chemical reactions or the rates of origination of new species in the biological studies). The results show that in the case with multiple singleton limit states , the rates are different and depend upon the smoothness of the urn function $f(.)$ in neighbourhoods of the states. It has occured that 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, other 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 in the previous chapter represent a powerful and convenient tool for studying and demonstrating the nonlinear effects pertinent to the processes.

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

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

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

Now designate $\zeta_i(f(x))$ by $\xi_i(x)$ and using the recursion (1.4) we can derive analogs of the above theorems for the generalized urn scheme.

4 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.

[3] Dosi, G., Yu. Ermoliev and Yu. Kaniovski (1991). *Generalized Urn Schemes and Technological Dynamics*, IIASA working paper WP-91-9. International Institute for Applied Systems Analysis, Laxenburg, Austria.

[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, **2**. pp. 136–138

[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. (1992). *When Are Touchpoints Limits for Generalized Polya Urns*, Proceedings of the American Mathematical Society (forthcomming).

[14] Robbins, H. and S. Monro (1951). *A Stochastic Approximation Method.* Ann. Math. Statist., **22**, pp. 400–407.

[15] Sacks, J. (1958). *Asymptotic Distribution of Stochastic Approximation Procedures.* Ann. Math. Stat. **29** pp. 375–405.

[16] Strassen, V. (1965). *The Existence of Probability Measures with Given Martingals*, Ann. Math. Statist., **36** pp. 423–439.