Exakte Abschätzungen für die Konvergenzgeschwindigkeit allgemeiner stochastischer Approximationsverfahren in $\mathbb{R}$

Valery Koval*
Hochschule für Gerätebau
Rusjkastraße 56
282001 Ternopil, Ukraine

Rainer Schwabe†
Freie Universität Berlin
1. Mathematisches Institut
Arnimallee 2–6
D-14 195 Berlin, Deutschland

Zusammenfassung

In dieser Arbeit werden für nichtlineare Differenzengleichungen der Form $x_n = x_{n-1} - a_n f(x_{n-1}) + b_n v_n$ exakte Raten und Konstanten für die Konvergenz von x_n gegen die Nullstelle θ der Funktion f unter sehr allgemeinen Voraussetzungen an die Schrittweiten a_n und die Gewichtungsfaktoren b_n der zufälligen Störungen, sowie an die Störungen v_n selbst, bestimmt und eine fast sichere Repräsentation der Abweichungen $x_n - \theta$ durch gewichtete Summen (unabhängiger) Zufallsgrößen hergeleitet.

1. Einleitung

Wir betrachten in $\mathbb{R}$ die Differenzengleichung

$$x_n = x_{n-1} - a_n f(x_{n-1}) + b_n v_n, \qquad n \geq 1,\ x_0 \in \mathbb{R}, \tag{1}$$

wobei $a_n > 0$, $\lim_{n\to\infty} a_n = 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $(b_n) \subset \mathbb{R}$, (v_n) eine zufällige Folge ist und die Gleichung $f(x) = 0$ die Lösung θ besitzt. Es gelte $x_n \to \theta$ f. s. $(n \to \infty)$. Wir wollen eine Folge von Konstanten (c_n) bestimmen, so daß

$$\limsup_{n\to\infty} c_n |x_n - \theta| = 1 \quad \text{f. s.} \tag{2}$$

---

*Diese Arbeit entstand während eines durch den Deutschen Akademischen Austauschdienstes geförderten Aufenthaltes an der Freien Universität Berlin

†gefördert durch die Forschungsprojekte Ku719/2-1 und Ku719/2-2 der Deutschen Forschungsgemeinschaft

gilt. Außerdem wollen wir die Lösung (x_n) der Gleichung (1) durch eine gewichtete Summe von zufälligen Größen (ξ_n) approximieren, d. h.

$$|x_n - \theta - \textstyle\sum_{k=1}^{n} d_{nk}\xi_k| = o(s_n) \quad \text{f. s.} \tag{3}$$

mit Konstanten (d_{nk}) und einer Folge von Konstanten (s_n) mit $s_n \to 0$ $(n \to \infty)$.

Diese Probleme wurden von verschiedenen Wissenschaftlern sowohl in $\mathbb{R}$ als auch in allgemeineren Räumen untersucht. Wir erwähnen hier nur die Arbeiten von Gaposhkin und Krasulina (1974), Korostelev (1983, 1984) für das Problem (2) und Kersting (1977), Ruppert (1982), Mark (1982) und Walk (1988, 1992) für beide Problemstellungen (2) und (3). In allen diesen Publikationen sahen die Folgen (a_n) und (b_n) im allgemeinsten Fall so aus: $a_n = n^{-a}$, $0 < a \leq 1$ und $b_n = n^{-b}$ mit $b \geq a/2$ (siehe z. B. Korostelev (1983, 1984) und Walk (1992)). Für die zufällige Folge (v_n) wurde üblicherweise vorausgesetzt, daß:

a) (v_n) eine Folge von unabhängigen, identisch verteilten zufälligen Größen ist (Korostelev (1983, 1984)),

b) (v_n) eine Martingaldifferenzfolge ist (Gaposhkin und Krasulina (1974)),

c) (v_n) einem log log-Invarianzprinzip genügt (Mark (1982), Walk (1988, 1992)),

und ähnliches (siehe z. B. Ruppert (1982)).

In der vorliegenden Arbeit werden die Problemstellungen (2) und (3) für Lösungen der Gleichung (1) unter hinreichend allgemeinen Voraussetzungen an (a_n), (b_n) und (v_n), die die weiter oben vorgestellten Voraussetzungen umfassen, untersucht. Zusätzlich wurde häufig für die Funktion $f(x)$ gefordert, daß $f(x) = \lambda(x - \theta) + O(|x - \theta|^{1+\delta})$ für $x \to \theta$ mit $\delta > 0$ (siehe z. B. Kersting (1977) und Ruppert (1982)). Wir benutzen hier die schwächere Bedingung $f(x) = \lambda(x - \theta) + o(x - \theta)$ für $x \to \theta$, d. h. f ist differenzierbar in θ. Wir werden auch die Beziehung (2) unter konkreteren Voraussetzungen an (v_n) untersuchen, um die Klasse der in Gleichung (1) zugelassenen Folgen (a_n) und (b_n) erweitern zu können.

2. Voraussetzungen

Wir betrachten die Gleichung (1) und geben Voraussetzungen an, die wir im weiteren benutzen werden. Wir setzen voraus, daß

$$f(x) = \lambda(x-\theta) + o(x-\theta), \qquad \text{für } x \to \theta, \text{ mit } \lambda > 0, \tag{4}$$

und

$$x_n \to \theta \quad \text{f. s. } (n \to \infty). \tag{5}$$

Bemerkung 1. Bedingung (4) ist schwächer als die übliche, die z. B. von Kersting (1977) und Ruppert (1982) benutzt wurde. Daher erhalten wir zwar eine schlechtere Genauigkeit der Approximation in (3), aber diese Approximation ist für die Lösung der Aufgabe (2) ausreichend (siehe auch Mark (1982), S. 59).

Bemerkung 2. Voraussetzung (5) wird üblicherweise gefordert bei Untersuchungen der Probleme (2) und (3) für die nichtlineare Gleichung (1) (siehe z. B. Ruppert (1982), Mark (1982) und Korostelev (1984)). Die Beziehung (5) wurde auch von Ljung (1978) und Li (1994) unter allgemeinen Voraussetzungen an die Folgen (a_n) und (b_n) betrachtet.

Für die Folgen (a_n) und (b_n) werden folgende Abschätzungen vorausgesetzt:

$$b_n^2/a_n \geq b_{n+1}^2/a_{n+1}, \qquad n \geq n_1 \geq 1, \tag{6}$$

und

$$b_{n+1}^2/a_{n+1} \geq (b_n^2/a_n)(1-\lambda a_{n+1})^{\nu}, \qquad n \geq n_2 \geq 1, \tag{7}$$

für ein $\nu \in [1,2)$.

Bemerkung 3. Bedingung (7) ist eine Verallgemeinerung der Bedingung $a_n \geq a_{n-1}(1-a_n)$, die von Ljung (1974) bei der Untersuchung der Gleichung $x_n = x_{n-1} - a_n(x_{n-1}+v_n)$ benutzt wurde. Die Bedingungen (6) und (7) wurden von Koval (1994) bei der Lösung der Aufgabe (2) für lineare Gleichungen (1) im $\mathbb{R}^d$ bei vorliegenden normalverteilten Störungen (v_n) verwendet.

Beispiel 1. Es sei $a_n = n^{-a}$, $0 < a < 1$, und $b_n = n^{-b}$ mit $a/2 < b \leq a$ (vgl. Korostelev (1984)). Dann sind die Bedingungen (6) und (7) erfüllt (sogar unter der abgeschwächten Voraussetzung $b > a/2$) mit $\nu = 1$.

Beispiel 2. Es sei $a_n = n^{-1}$ und $b_n = n^{-(1+\beta)/2}$ mit $\beta \geq 0$ und $\lambda > \beta/2$ (cgl. Walk (1992)). Dann gelten die Bedingungen (6) und (7). Wir möchten in diesem Zusammenhang anmerken, daß $\nu \geq \beta/\lambda$ gelten muß, damit die Bedingung (7) erfüllt ist. Insbesondere muß γ nahe bei 2 für λ nahe bei $\beta/2$ sein.

Beispiel 3. Es sei $a_n = b_n = 1/\log^{\alpha} n$, $\alpha > 0$, $n > 1$. Dann gelten die Bedingungen (6) und (7) mit $\nu = 1$.

Bemerkung 4. Aus Bedingung (7) folgt, daß $b_n \neq 0$ für alle $n \geq n_3$ ist, sofern $b_{n_3} \neq 0$ für $n_3 \geq 1$.

Wir setzen zusätzlich

$$\sup_n |b_n - b_{n+1}|(a_{n+1}b_n)^{-1} < \infty. \tag{8}$$

für die Folgen (a_n) und (b_n) voraus. Diese Bedingung kann auch umgeschrieben werden zu: $b_{n+1} \geq b_n(1 - ca_{n+1})$, $n \geq n_4$, für ein $c > 0$ (vgl. Bedingung (7)).

Beispiel 4. Für $a_n = n^{-a}$, mit $0 < a \leq 1$, und $b_n = n^{-b}$, mit $b > 0$, oder $a_n = b_n = 1/\log^{\alpha} n$, $n > 1$, mit $\alpha > 0$, ist Bedingung (8) erfüllt.

Wir betrachten als nächstes Voraussetzungen an die zufällige Folge (v_n). Ohne Beschränkung der Allgemeinheit können wir voraussetzen, daß

$$a_n < 1/\lambda, \qquad n \geq 1.$$

Weiterhin setzen wir

$$p_n \quad = \quad \prod_{j=1}^{n}(1 - \lambda a_j)\ ,$$

$$
\begin{aligned}
B_n &= \textstyle\sum_{k=1}^n b_k^2/p_k^2 \ , \\
\sigma_n^2 &= p_n^2 B_n \ , \\
L_n &= \log\log B_n \ , \\
\chi_n &= (2B_n L_n)^{1/2} \text{ und} \\
\ell_n &= \log(\textstyle\sum_{j=1}^n a_j) \ .
\end{aligned}
$$

Wir werden später zeigen, daß aus (6) und (7) $\lim_{n\to\infty} B_n = \infty$ folgt.

Bemerkung 5. Wir möchten anmerken, daß $B_n = Var(\sum_{k=1}^n (b_k/p_k)v_k)$, wenn (v_n) z. B. eine Folge von unkorrelierten zufälligen Größen mit $Ev_n^2 = 1$ ist. Alle Ergebnisse dieser Arbeit bleiben gültig, wenn wir statt (B_n) eine Folge (B_n') betrachten mit $B_n' \sim B_n$ $(n \to \infty)$.

Für (v_n) setzen wir insbesondere voraus, daß

$$\limsup_{n\to\infty} \chi_n^{-1} \left|\textstyle\sum_{k=1}^n (b_k/p_k)v_k\right| < \infty \qquad \text{f. s.} \tag{9}$$

Diese Voraussetzung ist erfüllt, wenn die Folge $(\sum_{k=1}^n (b_k/p_k)v_k)$ dem Gesetz vom iterierten Logarithmus genügt (siehe z. B. Stout (1974) und Petrov (1995)). Da die Summanden (b_n/p_n) eine sehr spezielle Gestalt besitzen, können wir später ein Kriterium für die Gültigkeit der Beziehung (9) und der spezielleren Beziehung (13) (siehe unten) unter hinreichend allgemeinen Voraussetzungen angeben (siehe Lemma 6 im Paragraph 4). Eine andere Bedingung an (v_n) wird in (16) gestellt.

Wir können ohne Einschränkung voraussetzen, daß $x_0 = 0$ ist, daß für die Lösung $\theta = 0$ gilt und daß $n_1 = n_2 = n_3 = 1$ gesetzt werden kann.

3. Hauptresultate

Satz 1. *Sind die Bedingungen (4) bis (7) und (9) erfüllt, dann gilt*

$$\lim_{n\to\infty} \left| x_n/(2\sigma_n^2 \ell_n)^{1/2} - \chi_n^{-1} \textstyle\sum_{k=1}^n (b_k/p_k)v_k \right| = 0 \qquad \textit{f. s.} \tag{10}$$

und

$$|x_n - p_n \textstyle\sum_{k=1}^n (b_k/p_k)v_k| = o\left((b_n^2 a_n^{-1} \ell_n)^{1/2}\right) \qquad \textit{f. s.} \tag{11}$$

Beziehung (11) stellt eine Lösung des Problems 3 dar (vgl. Ruppert (1982) für den Fall $a_n = n^{-1}$).

Bemerkung 6. Betrachten wir die lineare Differenzengleichung

$$y_n = y_{n-1} - a_n \lambda y_{n-1} + b_n v_n, \qquad n \geq 1, \ y_0 = 0, \tag{12}$$

dann gilt $y_n = p_n \sum_{k=1}^n (b_k/p_k) v_k$ und wir können die Beziehung (12) wie folgt umschreiben

$$|x_n - y_n| = o\left((b_n^2 a_n^{-1} \ell_n)^{1/2}\right) \qquad \text{f. s.}$$

Satz 2. *Sind die Bedingungen (4) bis (7) erfüllt und ist zusätzlich*

$$\limsup_{n\to\infty} \chi_n^{-1} \left|\sum_{k=1}^n (b_k/p_k) v_k\right| = 1 \qquad \textit{f. s.,} \tag{13}$$

dann gilt

$$\limsup_{n\to\infty} |x_n|/(2\sigma_n^2 \ell_n)^{1/2} = 1 \qquad \textit{f. s.} \tag{14}$$

Beispiel 5. Es sei $a_n = n^{-1}$, $b_n = n^{-(1+\beta)/2}$ mit $\beta \geq 0$, $\lambda > \beta/2$. Dann kann man nachrechnen, daß

$$\sigma_n^2 \sim n^{-\beta}/(2\lambda - \beta) \qquad (n \to \infty),$$

und

$$\ell_n \sim \log\log n \qquad (n \to \infty)$$

gilt. Somit erhalten wir aus (14)

$$\limsup_{n\to\infty} |x_n|/(n^{-\beta} \log\log n)^{1/2} = \left(\tfrac{2}{2\lambda-\beta}\right)^{1/2} \qquad \text{f. s.}$$

(Für den Fall $\beta = 1$ vgl. Gaposhkin und Krasulina (1974) und Kersting (1977), für $0 < \beta \leq 1$, $\lambda > 1/2$ vgl. Korostelev (1984) und Ruppert (1982), und im allgemeinen Fall siehe auch Walk (1992)). Wir möchten hier noch anmerken, daß $\beta > 0$ gelten muß, um $x_n \to 0$ f. s. zu sichern.

Beispiel 6. Es sei $a_n = n^{-a}$, $0 < a < 1$, und $b_n = n_{-b}$, $b > a/2$. Dann kann man berechnen, daß

$$\sigma_n^2 \sim n^{a-2b}/2\lambda \qquad (n \to \infty)$$

und

$$\ell_n \sim (1-a)\log n \qquad (n \to \infty)$$

gilt. Somit erhalten wir aus (14)

$$\limsup_{n\to\infty} |x_n|/(n^{a-2b}\log n)^{1/2} = \left(\tfrac{1-a}{\lambda}\right)^{1/2} \qquad \text{f. s.}$$

(Für den Fall $a/2 < b \leq a$ vgl. Korostelev (1984)).

Die Berechnung von σ_n^2 in (14) bereitet im allgemeinen Fall einige technische Schwierigkeiten. Jedoch reicht es in einigen Situationen aus, lediglich eine obere und untere Abschätzung für $\limsup_{n\to\infty} |x_n|/(2\sigma_2^n \ell_n)^{1/2}$ zu finden. Dazu kann man die folgende Behauptung benutzen:

Folgerung 1. *Sind die Bedingungen (4) bis (7) und (13) erfüllt, dann gilt*

$$1/\lambda^{1/2} \leq \limsup_{n\to\infty} |x_n|/(b_n^2 a_n^{-1} \ell_n)^{1/2} \leq \left(\frac{2}{\lambda(2-\nu)}\right)^{1/2} \qquad \textit{f. s.} \tag{15}$$

In der Beziehung (11) wird die Lösung (x_n) der Diffenrenzengleichung (1) durch eine gewichtete Summe der zufälligen Größen (v_n) approximiert. Im allgemeinen Fall können diese zufälligen Größen abhängig sein. Wir möchten daher die Lösung (x_n) auch durch eine gewichtete Summe von unabhängigen zufälligen Größen approximieren.

Satz 3. *Es seien die Bedingungen (4) bis (8) erfüllt und es existiere eine Folge von unabhängigen, standard-normalverteilten Zufallsfgrößen* (γ_n) *mit*

$$|\sum_{i=1}^n v_i - \sum_{i=1}^n \gamma_i| = o\left((a_n^{-1}\ell_n)^{1/2}\right) \qquad \textit{f. s.} \tag{16}$$

Dann gilt

$$|x_n - p_n \sum_{k=1}^n (b_k/p_k)\gamma_k| = o\left((b_n^2 a_n^{-1}\ell_n)^{1/2}\right) \qquad \textit{f. s.} \tag{17}$$

Sei die Folge (z_n) Lösung der linearen Differenzengleichung

$$z_n = (1-\lambda a_n)z_{n-1} + b_n\gamma_n, \qquad n \geq 1,\ z_0 = 0, \tag{18}$$

dann kann man (17) folgendermaßen umschreiben

$$|x_n - z_n| = o\left((b_n^2 a_n^{-1}\ell_n)^{1/2}\right) \qquad \text{f. s.} \tag{19}$$

Wir möchten hier anmerken, daß (z_n) wegen (18) ein zentrierter Gaußscher Markov-Prozeß mit diskreter Zeit ist.

Ähnliche Problemstellungen zu (17) und (19) wurden im Fall $a_n = n^{-1}$ von Kersting (1977) und Walk (1988, 1992) und im Fall $a_n = n^{-a}$, $0 < a < 1$, von Schwabe und Walk (1995) für die Lösung (y_n) der linearen Gleichung (12) untersucht.

Wir wollen jetzt die Bedingung (16) etwas genauer betrachten. Insbesondere möchten wir bemerken, daß diese Bedingung nur von der Folge (a_n) abhängt. Diese Bedingung sagt aus, daß wir strengere Anforderungen an die Genauigkeit Δ_n der Approximation der Summe $V_n = \sum_{i=1}^n v_i$ durch die Summe $\Gamma_n = \sum_{i=1}^n \gamma_i$ stellen müssen, wenn wir die Schrittweite (a_n) in der Gleichung (1) vergrößern, zum Beispiel:

a) für $a_n = n^{-1}$ muß $\Delta_n = o\left((\log\log n)^{1/2}\right)$

b) für $a_n = n^{-a}$, $0 < a < 1$, muß $\Delta_n = o\left((n^a \log n)^{1/2}\right)$

c) für $a_n = 1/\log^\alpha n)$, $\alpha > 1$ muß $\Delta_n = o\left((\log n)^{(1+\alpha)/2}\right)$

gelten. Aussagen über die Möglichkeit der Approximation von V_n durch Γ_n in den Fällen a), b) und c) kann man (unter verschiedenen Forderungen an die Folge (v_n)) z. B. bei Berger (1990), Philipp und Stout (1975), Komlós, Major und Tusnady

(1975, 1976) und Major (1976,1979) finden (siehe auch Bingham (1986) und Shao (1993)).

Satz 4. *Sind die Bedingungen (4) bis (8) und (16) erfüllt, dann gilt (14).*

Um die Klasse der zugelassenen Folgen (a_n) und (b_n) in Gleichung (1) zu erweitern, muß man konkretere Voraussetzungen an die statistischen Eigenschaften der Folge (v_n) machen. Wir betrachten zuerst Folgen von unabhängigen, identisch verteilten zufälligen Größen. In diesem Fall läßt sich wegen (13) das Problem (14) auf das Gesetz vom iterierten Logarithmus für gewichtete Summen von unabhängigen, identisch verteilten zufälligen Größen zurückführen (siehe z. B. Klesov (1987)).

Behauptung 1. *Es sei (v_n) eine Folge von unabhängigen, identisch verteilten zufälligen Größen mit $Ev_1 = 0$, $E|v_1|^{2/\gamma} < \infty$ für ein $\gamma \in (\frac{2}{3}, 1]$, die Bedingungen (4) bis (7) seien erfüllt und es gelte*

$$\sup_n a_n n^\gamma / \ell_n < \infty, \tag{20}$$

dann gilt (14).

Aus (20) folgt, daß man $a_n = n^{-1}$ wählen muß, wenn $\gamma = 1$ ist, d. h. wenn nur das zweite Moment Ev_1^2 existiert.

Wir möchten uns jetzt für den Fall interessieren, daß die Schrittweite (a_n) größer als n^{-1} gewählt werden kann.

Behauptung 2. *Es sei (v_n) eine Folge von unabhängigen, identisch verteilten zufälligen Größen mit $Ev_1 = 0$, $E|v_1|^{2+\delta} < \infty$ für ein $\delta > 0$, die Bedingungen (4) bis (7) seien erfüllt und es gelte*

$$\sup_n a_n \ell_n (\textstyle\sum_{j=1}^n a_j)^{2/\delta} < \infty, \tag{21}$$

dann gilt (14).

Bemerkung 7. Die Bedingungen (20) und (21) hängen, wie (16), nur von der Folge (a_n) ab.

Beispiel 7. Es sei $a_n = n^{-a}$, $0 < a < 1$, dann ist (21) für $\delta > 2(1-a)/a$ erfüllt. Wir bemerken noch, daß $\delta \uparrow \infty$ wenn $a \downarrow 0$.

Die Bedingung (21) ist z. B. dann nicht erfüllt, wenn $a_n = 1/\log^{\alpha} n$, $\alpha > 0$.

Wir möchten uns als nächstes dafür interessieren, was wir erhalten, wenn die zufälligen Größen (v_n) alle absoluten Momente besitzen, z. B. wenn sie beschränkte oder wenn sie normalverteilte zufällige Größen sind.

Behauptung 3. *Es sei (v_n) eine Folge von unabhängigen zufälligen Größen mit $Ev_n = 0$ und $|v_n| \le c < \infty$ f. s., die Bedingungen (4) bis (7) seien erfüllt und es gelte*

$$a_n \ell_n \to 0 \qquad (n \to \infty),$$

dann gilt (14).

Behauptung 4. *Es sei (v_n) eine Folge unabhängiger standard-normalverteilter zufälliger Größen und die Bedingungen (4) bis (7) seien erfüllt, dann gilt (14).*

Beispiel 8. Es sei $a_n = b_n = 1/\log^{\alpha} n$, $\alpha > 0$. Dann sind die Voraussetzungen (6) und (7) erfüllt und wir erhalten (unter Ausnutzung der Ungleichung (15))

$$\lambda^{-1/2} \le \limsup_{n\to\infty} |x_n|(\log n)^{(\alpha-1)/2} \le (2/\lambda)^{1/2} \qquad \text{f. s.}$$

Wir möchten hier noch anmerken, daß $\alpha > 1$ sein muß, um $x_n \to 0$ f. s. zu sichern.

4. Hilfsresultate

Lemma 1. *Unter den Voraussetzungen (6) und (7) gilt*

$$\tfrac{1}{2\lambda}\,\tfrac{b_n^2}{a_n}(1+o(1)) \le \sigma_n^2 \le \tfrac{1}{(2-\nu)\lambda}\,\tfrac{b_n^2}{a_n}.$$

Beweis. Wir beweisen zuerst, daß

$$\sigma_n^2 \geq (2\lambda)^{-1}\,(1+o(1))\,b_n^2/a_n. \tag{22}$$

Dazu benutzen wir die Bedingung (6). Wir vereinbaren $\prod_{j=n+1}^n (1-\lambda a_j) = 1$.

$$\begin{aligned}
\sigma_n^2 &= \textstyle\sum_{k=1}^n b_k^2 \prod_{j=k+1}^n (1-\lambda a_j)^2 \\
&= \textstyle\sum_{k=1}^n (b_k^2/a_k) a_k \prod_{j=k+1}^n (1-\lambda a_j)^2 \\
&\geq (b_k^2/a_n) \textstyle\sum_{k=1}^n \frac{a_k}{1-(1-\lambda a_k)^2} \left(\prod_{j=k+1}^n (1-\lambda a_j)^2 - \prod_{j=k}^n (1-\lambda a_j)^2\right) \\
&\geq (b_n^2/a_n)(2\lambda)^{-1} \textstyle\sum_{k=1}^n \left(\prod_{j=k+1}^n (1-\lambda a_j)^2 - \prod_{j=k}^n (1-\lambda a_j)^2\right) \\
&= (2\lambda)^{-1}(b_n^2/a_n)\left(1-\prod_{j=1}^n (1-\lambda a_j)^2\right) \ .
\end{aligned}$$

Daraus folgt (22), da $\prod_{j=1}^n (1-\lambda a_j)^2 \to 0$ $(n \to \infty)$ auf Grund der Voraussetzung $\sum_{j=1}^\infty a_j = \infty$.

Wir beweisen jetzt, daß

$$\sigma_n^2 \leq ((2-\nu)\lambda)^{-1}\, b_n^2/a_n$$

gilt. Dazu benutzen wir

$$\textstyle\prod_{j=k+1}^n (1-\lambda a_j)^\nu b_k^2/a_k \leq b_n^2/a_n, \qquad 1 \leq k \leq n,$$

was durch Iteration aus Bedingung (7) folgt. Damit gilt

$$\begin{aligned}
\sigma_n^2 &= \textstyle\sum_{k=1}^n \left(\prod_{j=k+1}^n (1-\lambda a_j)^\nu b_k^2/a_k\right) a_k \prod_{j=k+1}^n (1-\lambda a_j)^{2-\nu} \\
&\leq (b_n^2/a_n) \textstyle\sum_{k=1}^n \frac{a_k}{1-(1-\lambda a_k)^{2-\nu}} \left(\prod_{j=k+1}^n (1-\lambda a_j)^{2-\nu} - \prod_{j=k}^n (1-\lambda a_j)^{2-\nu}\right) \\
&\leq ((2-\nu)\lambda)^{-1}\,(b_n^2/a_n) \textstyle\sum_{k=1}^n \left(\prod_{j=k+1}^n (1-\lambda a_j)^{2-\nu} - \prod_{j=k}^n (1-\lambda a_j)^{2-\nu}\right) \\
&= ((2-\nu)\lambda)^{-1}\,(b_n^2/a_n)\left(1-\prod_{j=1}^n (1-\lambda a_j)^{2-\nu}\right) \\
&\leq ((2-\nu)\lambda)^{-1}\,(b_n^2/a_n)\ ,
\end{aligned}$$

für alle $n \geq 1$. □

Bemerkung 8. In manchen Situationen ist es ausreichend, lediglich eine Abschätzung $\sigma_n^2 \geq (a_1/2)b_n^2/a_n$, für alle $n \geq 1$, zu erhalten.

Lemma 2. *Unter den Bedingungen (6) und (7) gilt* $B_n \to \infty$ $(n \to \infty)$.

Beweis. Es gilt

$$\begin{aligned} B_n &= \sigma_n^2/p_n^2 \\ &\geq (a_1/2)b_n^2 a_n^{-1} \textstyle\prod_{j=1}^n (1-\lambda a_j)^{-\nu} \prod_{j=1}^n (1-\lambda a_j)^{\nu-2} \\ &\geq \left(4\lambda(1-\lambda a_1)^\nu\right)^{-1} b_1^2 \textstyle\prod_{j=1}^n (1-\lambda a_j)^{\nu-2} \,. \end{aligned} \tag{23}$$

Daraus folgt die Behauptung des Lemmas, da $\nu - 2 < 0$ und $\sum_{j=1}^\infty a_j = \infty$. □

Im weiteren bezeichnen wir mit $M,\ M_1,\ M_2,\ \ldots$ geeignete positive Konstanten.

Lemma 3. *Unter den Bedingungen (6) und (7) gilt* $L_n \sim \ell_n$ $(n \to \infty)$.

Beweis. Nach Definiton gilt $L_n = \log\log B_n = \log\log(\sigma_n^2/p_n^2)$. Weiterhin gilt einerseits (siehe (23))

$$B_n \geq M_1 \textstyle\prod_{j=1}^n (1-\lambda a_j)^{\nu-2} \geq M_1 \exp\left((2-\nu)\lambda \sum_{j=1}^n a_j\right)$$

und andererseits (siehe Lemma 1 und (6))

$$B_n \leq M_2 b_n^2 a_n^{-1} p_n^{-2} \leq M_2 a_1^{-1} b_1^2 \textstyle\prod_{j=1}^n (1-\lambda a_j)^{-2} \leq M_3 \exp\left((2\lambda+\varepsilon)\sum_{j=1}^n a_j\right) \,,$$

für ein $\varepsilon > 0$, da $\lim_{j\to\infty} a_j = 0$. Somit ist für große n

$$L_n' := \log(\log M_1 + (2-\nu)\lambda \textstyle\sum_{j=1}^n a_j) \leq L_n \leq \log(\log M_3 + (2\lambda+\varepsilon)\sum_{j=1}^n a_j) =: L_n''.$$

Wegen $L_n' \sim \ell_n$ und $L_n'' \sim \ell_n$ $(n \to \infty)$ folgt die Behauptung des Lemmas. □

Lemma 4. *Ist (z_n) Lösung der Differenzengleichung (18), dann gilt unter den Voraussetzungen (6) und (7)*

$$\limsup_{n\to\infty} |z_n|/(2\sigma_n^2\ell_n)^{1/2} = 1 \qquad f.\,s.$$

Beweis. Wir schreiben z_n als $z_n = p_n\tilde{z}_n$, wobei $\tilde{z}_n = \sum_{k=1}^n (b_k/p_k)\gamma_k$. Dann gilt $\tilde{z}_n = W(B_n)$, wobei $W(\cdot)$ eine standardisierte Brownsche Bewegung ist. Weiter gilt $B_n \to \infty$ $(n\to\infty)$ auf Grund von Lemma 2.

Wegen Bedingung (6) gilt $1 \le B_{n+1}/B_n = 1 + b_{n+1}^2 \prod_{j=1}^{n+1}(1-\lambda a_j)^{-2} B_n^{-1} = 1 + b_{n+1}^2(1-\lambda a_{n+1})^{-2}\sigma_n^{-2} \le 1 + M a_n b_{n+1}^2 b_n^2 \le 1 + M a_{n+1}$, woraus mit $\lim_{n\to\infty} a_n = 0$

$$\lim_{n\to\infty} B_{n+1}/B_n = 1 \tag{24}$$

für das Wachstum von B_n folgt.

Daher folgt aus dem Gesetz vom iterierten Logarithmus für die Brownsche Bewegung mit Zeit $t = B_n$ (siehe z. B. Gut (1986))

$$\limsup_{n\to\infty} |\tilde{z}_n|/(2B_n \log\log B_n)^{1/2} = 1 \qquad \text{f. s.}$$

bzw.

$$\limsup_{n\to\infty} |z_n|/(2\sigma_n^2 L_n)^{1/2} = 1 \qquad \text{f. s.}$$

Aus dieser Beziehung und Lemma 3 folgt die Behauptung von Lemma 4. □

Lemma 5. *Es seien (y_n) und (z_n) Lösungen der Differenzengleichungen (12) bzw. (18). Sind die Bedingungen (7), (8) und (16) erfüllt, dann gilt*

$$|y_n - z_n| = o\left((b_n^2 a_n^{-1}\ell_n)^{1/2}\right) \qquad f.\,s. \tag{25}$$

Beweis. Wir setzen $V_n = \sum_{i=1}^n v_i$, $\Gamma_n = \sum_{i=1}^n \gamma_i$ und benutzen die Abelsche Identität:

$$y_n - z_n \;=\; \sum_{k=1}^n b_k \prod_{j=k+1}^n (1-\lambda a_j)(v_k - \gamma_k)$$

$$
\begin{aligned}
&= \sum_{k=1}^{n-1} \left(b_k \prod_{j=k+1}^{n} (1-\lambda a_j) - b_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j) \right) (V_k - \Gamma_k) + b_n (V_n - \Gamma_n) \\
&= \sum_{k=1}^{n-1} \prod_{j=k+2}^{n} (1-\lambda a_j) b_k a_{k+1} \left(\frac{b_k - b_{k+1}}{a_{k+1} b_k} - \lambda \right) (\ell_k / a_k)^{1/2} \left((V_k - \Gamma_k)(a_k/\ell_k)^{1/2} \right) \\
&\quad + b_n (V_n - \Gamma_n) \ .
\end{aligned}
$$

Wir setzen

$$\delta_n = |V_n - \Gamma_n| (a_n/\ell_n)^{1/2}.$$

Auf Grund der Bedingung (16) gilt

$$\delta_n \to 0 \quad \text{f.s.} \qquad (n \to \infty). \tag{26}$$

Im weiteren benutzen wir (8) und die Bedingung

$$|b_k| a_k^{-1/2} \prod_{j=k+2}^{n} (1-\lambda a_j)^{\nu/2} \leq M_1 |b_n| a_n^{-1/2},$$

die aus (7) folgt, und erhalten

$$
\begin{aligned}
|y_n - z_n| &\leq \sum_{k=1}^{n-1} |b_k| a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j) \left(\left| \frac{b_k - b_{k+1}}{a_{k+1} b_k} \right| + \lambda \right) (\ell_k / a_k)^{1/2} \delta_k \\
&\quad + (b_n^2 a_n^{-1} \ell_n)^{1/2} \delta_n \\
&\leq M_2 \ell_n^{1/2} \sum_{k=1}^{n} \left(|b_k| a_k^{-1/2} \prod_{j=k+2}^{n} (1-\lambda a_j)^{\nu/2} \right) a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j)^{1-(\nu/2)} \delta_k \\
&\quad + (b_n^2 a_n^{-1} \ell_n)^{1/2} \delta_n \\
&\leq M_3 (b_n^2 a_n^{-1} \ell_n)^{1/2} \left(\sum_{k=1}^{n-1} a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j)^{1-(\nu/2)} \delta_k + \delta_n \right) \ .
\end{aligned}
$$

Also gilt insgesamt

$$|y_n - z_n| \leq M_3 (b_n^2 a_n^{-1} \ell_n)^{1/2} (T_n + \delta_n) \tag{27}$$

mit

$$T_n = \sum_{k=1}^{n-1} a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j)^{1-(\nu/2)} \delta_k \ .$$

Wir zeigen als nächstes, daß $T_n \to 0$ f.s. $(n \to \infty)$. Dazu setzen wir

$$c_{nk} = a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j)^{\alpha} \ , \qquad 1 \leq k \leq n-1.$$

Hier ist $\alpha = 1 - (\nu/2)$ und folglich $\alpha \in (0, 1/2]$. T_n läßt sich schreiben als

$$T_n = \sum_{k=1}^{n-1} c_{nk} \delta_k.$$

Insbesondere erhalten wir $c_{nk} \geq 0$ und

$$c_{nk} \to 0 \qquad (n \to \infty) \text{ für jedes } k \geq 1. \tag{28}$$

Weiterhin gilt

$$\begin{aligned}
&\textstyle\sum_{k=1}^{n-1} c_{nk} \sum_{k=1}^{n-1} \frac{a_{k+1}}{1-(1-\lambda a_{k+1})^\alpha} \left(\prod_{j=k+2}^{n}(1-\lambda a_j)^\alpha - \prod_{j=k+1}^{n}(1-\lambda a_j)^\alpha\right) \\
&\leq \; (\lambda\alpha)^{-1} \textstyle\sum_{k=1}^{n-1} \left(\prod_{j=k+2}^{n}(1-\lambda a_j)^\alpha - \prod_{j=k+1}^{n}(1-\lambda a_j)^\alpha\right) \\
&= \; (\lambda\alpha)^{-1} \left(1 - \prod_{j=2}^{n}(1-\lambda a_j)^\alpha\right) \\
&\leq \; (\lambda\alpha)^{-1} < \infty \, .
\end{aligned} \tag{29}$$

Aus den Beziehungen (26), (28) und (29) und dem Lemma von Toeplitz folgt

$$T_n \to 0 \quad \text{f. s.} \qquad (n \to \infty)$$

Hieraus und aus (26) folgt die Behauptung des Lemmas mit Hilfe der Ungleichung (27). □

Lemma 6. *Sind die Bedingungen (6) bis (8) und (16) erfüllt, dann gilt (13).*

Beweis. Auf Grund von Lemma 3 und 4 sowie 1 und 5 erhalten wir

$$\begin{aligned}
&\limsup_{n\to\infty} \left|\chi_n^{-1} \textstyle\sum_{k=1}^{n}(b_k/p_k)v_k \; - \; z_n/(2\sigma_n^2 \ell_n)^{1/2}\right| \\
&\leq \; \limsup_{n\to\infty} |y_n - z_n|/(2\sigma_n^2\ell_n)^{1/2} \\
&\leq \; \lambda^{1/2} \limsup_{n\to\infty} |y_n - z_n|/(b_n^2 a_n^{-1}\ell_n)^{1/2} = 0 \quad \text{f. s.}
\end{aligned}$$

Aus dieser Beziehung, Lemma 4 und der Dreiecks-Ungleichung

$$\left|\limsup_{n\to\infty} |s_n'| - \limsup_{n\to\infty} |s_n''|\right| \leq \limsup_{n\to\infty} |s_n' - s_n''|$$

erhalten wir

$$\limsup_{n\to\infty} \chi_n^{-1} \left|\sum_{k=1}^{n}(b_k/p_k)v_k\right| = \limsup_{n\to\infty} |z_n|/(2\sigma_n^2\ell_n)^{1/2} = 1 \quad \text{f. s.}$$

□

5. Beweise der Hauptresultate

Beweis des Satzes 1. Aus den Bedingungen (4) und (5) folgt

$$f(x_n) = \lambda x_n + \delta_n x_n,$$

wobei

$$\delta_n \to 0 \quad \text{f. s.} \qquad (n \to \infty).$$

Dann ist

$$x_n = (1 - \lambda a_n) x_{n-1} + a_n \delta_{n-1} x_{n-1} + b_n v_n, \quad n \geq 1,$$

und

$$x_n / p_n = \sum_{k=1}^{n} (b_k / p_k) v_k + \sum_{k=1}^{n-1} (a_{k+1} / p_{k+1}) \delta_k x_k.$$

Wir teilen die letzte Zeile durch $(B_n \ell_n)^{1/2}$:

$$x_n / (\sigma_n^2 \ell_n)^{1/2} = \sum_{k=1}^{n} (b_k / p_k) v_k / (B_n \ell_n)^{1/2} + \sum_{k=1}^{n-1} \frac{a_{k+1}}{1-\lambda a_{k+1}} \left(\frac{B_k \ell_k}{B_n \ell_n}\right)^{1/2} \delta_k x_k / (\sigma_k^2 \ell_k)^{1/2}$$

und setzen

$$\tilde{x}_n = x_n / (\sigma_n^2 \ell_n)^{1/2},$$

sowie

$$\begin{aligned} R_n &= \sum_{k=1}^{n} (b_k / p_k) v_k / (B_n \ell_n)^{1/2} , \\ G_{n,k} &= \frac{a_{k+1}}{1-\lambda a_{k+1}} \left(\frac{B_k \ell_k}{B_n \ell_n}\right)^{1/2} \delta_k, \quad 1 \leq k \leq n-1. \end{aligned}$$

Dann läßt sich $\tilde{x}_n$ darstellen als

$$\tilde{x}_n = R_n + \sum_{k=1}^{n-1} G_{n,k} \tilde{x}_k \ .$$

Zuerst wollen wir zeigen, daß

$$\sup_n |\tilde{x}_n| < \infty \quad \text{f. s.}$$

gilt. Wegen $(2 B_n \ell_n)^{1/2} \sim \chi_n$ (Lemma 3), folgt aus Bedingung (9)

$$\sup_n |R_n| < \infty \quad \text{f. s.} \tag{30}$$

Als nächstes wollen wir zeigen, daß

$$\lim_{n\to\infty} \sum_{k=1}^{n-1} |G_{n,k}| = 0 \qquad \text{f. s.} \tag{31}$$

gilt. Dazu schätzen wir wie folgt ab

$$\begin{aligned} \sum_{k=1}^{n-1} |G_{n,k}| &\leq \sum_{k=1}^{n-1} \frac{a_{k+1}}{1-\lambda a_{k+1}} \frac{p_n}{p_k} \frac{\sigma_k}{\sigma_n} |\delta_k| \\ &\leq M \sum_{k=1}^{n-1} \frac{a_{k+1}}{1-\lambda a_{k+1}} \frac{p_n}{p_k} \frac{|b_k|}{|b_n|} \frac{a_n^{1/2}}{a_k^{1/2}} |\delta_k| \\ &= M \left(a_n^{1/2}/|b_n|\right) \sum_{k=1}^{n-1} a_{k+1} \prod_{j=k+2}^{n} (1-\lambda a_j) \left(|b_k|/a_k^{1/2}\right) |\delta_k| \; . \end{aligned}$$

Weiter verfahren wir wie im Beweis von Lemma 5 und erhalten die Beziehung (31). Aus (30), (31) und Lemma 2 von Walk und Zsidó (1989) folgt

$$\sup_n |\tilde{x}_n| < \infty \qquad \text{f. s.}$$

Damit gilt

$$|\tilde{x}_n - R_n| \leq \sum_{k=1}^{n-1} |G_{n,k}| \, \sup_n |\tilde{x}_n| \; ,$$

woraus mit (31) die Beziehung (10) folgt. Beziehung (11) ergibt sich schließlich aus (10) und Lemma 1. □

Der **Beweis des Satzes 2** folgt unmittelbar aus den Beziehungen (10) und (13).

Beweis des Satzes 3. Unter den Bedingungen des Satzes gelten Lemma 5 und 6 sowie Satz 1. Damit folgt die Behauptung (17) aus den Beziehungen (11) und (25) (vgl. Bemerkung 6 und die Definition (18) von z_n). □

Der **Beweis des Satzes 4** folgt aus Beziehung (17) (Satz 3) und Lemma 4 bzw. aus Lemma 6 und Satz 2.

Der **Beweis der Folgerung 1** folgt aus Beziehung (14) (Satz 2) und Lemma 1.

Die **Beweise der Behauptungen 1 und 2** werden auf die Überprüfung der Bedingungen entsprechender Ergebnisse (Folgerung 4 und Beziehung (6)) in der Arbeit von Klesov (1987) zurückgeführt. (Dabei muß für Folgerung 4 die Beziehung (24) ausgenutzt werden.)

Der **Beweis der Behauptung 3** kann auf die Überprüfung der Bedingungen des klassischen Gesetzes vom iterierten Logarithmus von Kolmogorov (siehe z. B. Petrov (1995)) zurückgeführt werden.

Der **Beweis der Behauptung 4** folgt aus Satz 2 und Lemma 4, da in diesem Fall $v_n = \gamma_n$, $n \geq 1$, in Gleichung (1) gesetzt werden kann.

Literatur

[1] Berger, E. (1990). An almost sure invariance principle for stationary ergodic sequences of Banach space valued random variables. *Probab. Theory Related Fields* **84** 161–201.

[2] Bingham, N. H. (1986). Variants on the law of the iterated logarithm. *Bull. London Math. Soc.* **18** 433–467.

[3] Gaposhkin, V. F. und Krasulina, T. P. (1974). On the law of the iterated logarithm in stochastic approximation processes. *Theory Probab. Appl.* **19** 844–850.

[4] Gut, A. (1986). The law of the iterated logarithm for subsequences. *Probab. Math. Statist.* **7** 27–58.

[5] Kersting, G. (1977). Almost sure approximation of the Robbins-Monro process by sums of independent random variables. *Ann. Probab.* **5** 954–965.

[6] Klesov, O. J. (1987). The law of the iterated logarithm for weighted sums of independent identically distributed random variables. *Theory Probab. Appl.* **31** 337–342.

[7] Komlós, J., Major, P. und Tusnády, G. (1975, 1976). An Approximation of partial sums of independent RV's, and the sample DF, I, II. *Z. Wahrscheinlichkeitstheorie*

verw. Geb. **32** 111–131, **34** 33–58.

[8] Korostelev, A. P. (1983). On upper functions for stochastic approximation procedures. *Theory Probab. Appl.* **28** 806–811.

[9] Korostelev, A. P. (1984). *Stochastic recursive procedures (local properties).* Nauka, Moskau (auf russisch).

[10] Koval, V. A. (1994). On the rate of convergence for stochastic approximation procedures. *Ukrain. Math. Journ.* **46** 997–1002 (auf russisch).

[11] Li, G. (1994). Almost sure convergence of stochastic approximation procedures. *Statist. Sinica* **4** 361–372.

[12] Ljung, L. (1974). Convergence of recursive stochastic algorithms. *Report 7403, Dept. of Automatic Control, Lund Institute of Technology.*

[13] Ljung, L. (1978). Strong convergence of a stochastic approximation algorithm. *Ann. Statist.* **6** 680–696.

[14] Major, P. (1976). The approximation of partial sums of independent rv's. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **35** 213–220.

[15] Major, P. (1979). An improvement of Strassen's invariance principle. *Ann. Probab.* **7** 55–61.

[16] Mark, G.(1982). Loglog-Invarianzprinzipien für Prozesse der stochastischen Approximation. *Mitteilungen Math. Sem. Gießen* **153**.

[17] Petrov, V. V. (1995). *Limit theorems of probability theory: sequences of independent random variables.* Clarendon, Oxford.

[18] Philipp, W. und Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Am. Math. Soc.* **161**.

[19] Ruppert, D. (1982). Almost sure approximation to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Probab.* **10** 178–187.

[20] Schwabe, R. und Walk, H. (1995). On a stochastic approximation procedure based on averaging. *Metrika* (im Druck).

[20] Shao, Q.-M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Processes Appl.* **48** 319–334.

[21] Stout, W. (1974). *Almost sure convergence.* Academic Press, New York.

[22] Walk, H. (1988). Limit behaviour of stochastic approximation processes. *Statist. Decis.* **6** 109–128.

[23] Walk, H. (1992). Foundations of stochastic approximation. In: L. Ljung, G. Pflug und H. Walk: *Stochastic Approximation and Optimization of Random Systems, DMV Seminar Blaubeuren, May 28-June 4, 1989. DMV Seminar* **17**. Birkhäuser, Basel, 1–51.

[24] Walk, H. und Zsidó, L. (1989). Convergence of the Robbins-Monro method for linear problems in a Banach space. *J. Math. Anal. Appl.* **139** 152–177.