ISSN 1066-5307, Mathematical Methods of Statistics, 2008, Vol. 17, No. 2, pp. 138–145. © Allerton Press, Inc., 2008.

Asymptotic Expansions of the Robbins–Monro Process

J. Dippon[1*]

[1] *Institut für Stochastik und Anwendungen, Univ. Stuttgart 70550 Stuttgart, Germany*

Received July 5, 2007; in final form, March 19, 2008

Abstract—Assume that the function values $f(x)$ of an unknown regression function $f\colon \mathbb{R} \to \mathbb{R}$ can be observed with some random error V. To estimate the zero ϑ of f, Robbins and Monro suggested to run the recursion $X_{n+1} = X_n - \frac{a}{n} Y_n$ with $Y_n = f(X_n) - V_n$. Under regularity assumptions, the normalized Robbins–Monro process, given by $(X_{n+1} - \vartheta)/\sqrt{\mathrm{Var}(X_{n+1})}$, is asymptotically standard normal. In this paper Edgeworth expansions are presented which provide approximations of the distribution function up to an error of order $o(1/\sqrt{n})$ or even $o(1/n)$. As corollaries asymptotic confidence intervals for the unknown parameter ϑ are obtained with coverage probability errors of order $O(1/n)$. Further results concern Cornish–Fisher expansions of the quantile function, an Edgeworth correction of the distribution function and a stochastic expansion in terms of a bivariate polynomial in $1/\sqrt{n}$ and a standard normal random variable. The proofs of this paper heavily rely on recently published results on Edgeworth expansions for approximations of the Robbins–Monro process.

Key words: stochastic approximation, Robbins–Monro procedure, Edgeworth expansion.

2000 Mathematics Subject Classification: 62L20.

DOI: 10.3103/S106653070802004X

1. INTRODUCTION

This paper continues the investigation of Edgeworth expansions for the Robbins–Monro (R–M) process as started in [7]. In that paper we obtained Edgeworth expansions for sequences approximating the R–M process, which are easier to handle than the R–M process itself. Now we show that under appropriate conditions both processes possess the same Edgeworth expansions. Furthermore, several applications of these expansions are given.

To recall the definition of the R–M process assume that ϑ is the zero of an unknown regression function $f\colon \mathbb{R} \to \mathbb{R}$, whose values $f(x)$ at x can be observed with some random error only. In 1951 Robbins and Monro [13] suggested estimation of ϑ by X_n generated by the iteration

$$X_{n+1} = X_n - \frac{a}{n} Y_n.$$

Here $Y_n = f(X_n) - V_n$ is the noisy observation of f at X_n, V_n is the observation error, and $a > 0$ is some fixed number. As a result they obtained $X_n - \vartheta \to 0$ in probability ($n \to \infty$). Under the condition $af'(\vartheta) > 1/2$ Sacks [15] showed asymptotic normality of the sequence $((X_{n+1} - \vartheta)/\sigma_n)$, where $\sigma_n^2 := \mathrm{var}(X_{n+1})^2$ is of order $1/n$ asymptotically. Assuming $af'(\vartheta) > 1$ Renz [12] proved a limit theorem of Berry–Esseen type, which implies

$$P\left(\frac{X_{n+1} - \vartheta}{\sigma_n} \le x\right) = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right).$$

*E-mail: dippon@mathematik.uni-stuttgart.de

In this paper it will be shown that this result can be refined using the technique of Edgeworth expansions. Provided that $af'(\vartheta) > 1$ or $af'(\vartheta) > 3/2$ it is possible to obtain

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \leq x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + o\left(\frac{1}{\sqrt{n}}\right)$$

and

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \leq x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \frac{1}{n} p_2(x)\phi(x) + o\left(\frac{1}{n}\right)$$

respectively, with distribution function Φ and density ϕ of the standard normal distribution and some polynomials p_1, p_2 (Theorems 3 and 4).

These Edgeworth expansions can be used as starting points for other interesting expansions such as expansions of the coverage probability of confidence intervals (Corollaries 6 and 7), Cornish–Fisher expansions for the quantile function (Proposition 8) or for the Edgeworth correction of the distribution function of $\sqrt{n}(X_{n+1}-\vartheta)$ (Proposition 10). Besides the distribution function, the R–M process $(X_{n+1}-\vartheta)$ itself can be expanded as a sum of powers of $1/\sqrt{n}$ and a normally distributed random variable (Proposition 12).

Some Notation. For real numbers x and y their maximum is denoted by $x \vee y$. The standard deviation of a random variable X is abbreviated by $\sigma(X)$.

2. ASYMPTOTIC EXPANSIONS OF THE ROBBINS–MONRO PROCESS

We require the following two conditions.

Condition 1 (The Robbins–Monro process). *The regression function $f\colon \mathbb{R} \to \mathbb{R}$ is measurable; $V, V_1, V_2, \ldots$ is a sequence of i.i.d. real random variables satisfying $EV = 0$, $EV^2 = \sigma^2$, $EV^3 = \rho^3$, $EV^4 = \mu^4$ and, for some $m \geq 6$ (to be specified later on) $E|V|^m < \infty$; X_1 is a real random variable (the starting value) with $E|X_1|^m < \infty$. For a given $a > 0$ the recursion*

$$X_{n+1} = X_n - \frac{a}{n}(f(X_n) - V_n), \qquad n \in \mathbb{N},$$

defines the stochastic process (X_n).

Condition 2 (Quasi-linearity and local smoothness). *The function $f\colon \mathbb{R} \to \mathbb{R}$ is quasi-linear around ϑ in the sense that*

$$\underset{\vartheta\in\mathbb{R}}{\exists} \quad \underset{0<K_1<K_2<\infty}{\exists} \quad \underset{x\in\mathbb{R}}{\forall} \quad K_1|x-\vartheta| \leq \operatorname{sgn}(x-\vartheta)\cdot f(x) \leq K_2|x-\vartheta|.$$

Moreover, f is β-smooth at $\vartheta \in \mathbb{R}$ in the sense that the β_*th derivative of f is Hölder continuous of order $\beta - \beta_*$ at ϑ, where $\beta_* := \max\{n \in \mathbb{N}\colon n < \beta\}$. For brevity we will use $A = f'(\vartheta)$, $B = f''(\vartheta)$, and $C = f'''(\vartheta)$, whenever the relevant derivative exists.

Instead of asking for Hölder continuity of the derivative $f^{(\beta_*)}$ at ϑ it would be sufficient to require a little bit less, namely, the existence of some constants $c_1, \ldots, c_{\beta_*}$ satisfying

$$\left| f(x) - \sum_{i=1}^{\beta_*} \frac{1}{i!}\, c_i\, (x-\vartheta)^i \right| = O\big(|x-\vartheta|^\beta\big).$$

The graph of a quasi-linear function is enclosed between two straight lines intersecting at $(x, y) = (\vartheta, 0)$ and having positive slope. However, the following two theorems are even valid under the weaker assumption of sub-linearity [4].

Theorem 3. *Assume that Conditions* 1 *and* 2 *hold with* $aA > 1$, $aK_1 > 1/2$, $\beta > 2$ *and* $E|V|^m < \infty$ *for some* $m \geq 6$ *with* $m > 2(2aA-1)/(aA-1)$. *Then*

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \leq x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + o\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in $x \in \mathbb{R}$, *where* $\sigma_n^2 = \mathrm{var}(X_{n+1})$ *and*

$$p_1(x) = -\left(k_{1,2} + \frac{1}{6} k_{3,1}(x^2-1)\right)$$

with

$$k_{1,2} = -\frac{a^2 B\sigma}{(aA-1)\sqrt{2aA-1}} \qquad \text{and} \qquad k_{3,1} = \frac{(2aA-1)^{3/2}}{3aA-2}\frac{\rho^3}{\sigma^3} - \frac{a^2 B\sigma}{\sqrt{2aA-1}(3aA-2)}.$$

Theorem 4. *Assume that Conditions* 1 *and* 2 *hold with* $aA > 3/2$, $aK_1 > 1/2$, $\beta > 3$ *and* $E|V|^m < \infty$ *for some* $m \geq 12$ *with* $m > 3(2aA-1)/(aA-3/2)$. *Then*

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \leq x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \frac{1}{n} p_2(x)\phi(x) + o\left(\frac{1}{n}\right) \tag{1}$$

uniformly in $x \in \mathbb{R}$, *where*

$$p_2(x) = -x\left(\frac{1}{2}k_{1,2}^2 + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2-3) + \frac{1}{72}k_{3,1}^2(x^4 - 10x^2 + 15)\right)$$

with p_1, $k_{1,2}$ *and* $k_{3,1}$ *as given in Theorem* 3 *and*

$$k_{4,1} = \frac{(2aA-1)^2}{4aA-3}\frac{\mu^4}{\sigma^4} - 3\frac{4a^2A^2-1}{4aA-3} + \frac{36a^4B^2\sigma^2}{(2aA-1)(3aA-2)(4aA-3)} - \frac{12a^2(2aA-1)B}{(3aA-2)(4aA-3)}\frac{\rho^3}{\sigma^2} - \frac{4a^3C\sigma^2}{(2aA-1)(4aA-3)}.$$

Remark 5. (a) *The regularity assumption on* aA. Using step lengths $a_n = a/n$ the condition $aA > 1/2$ is necessary in order to obtain that the sequence $(\sqrt{n}(X_{n+1}-\vartheta))$ is asymptotically $N(0, a^2\sigma^2/(2aA-1))$-distributed. The only mimimum of the variance of the asymptotic distribution is attained for $aA = 1$. Since in applications $A = f'(\vartheta)$ is usually unknown, it was proposed to consider adaptive procedures which estimate A alongside the iteration [16]. Another possibility to circumvent this difficulty was suggested by Polyak [10] and Ruppert [14]. Instead of step lengths $a_n = a/n$ they apply more slowly decreasing step lengths such as $a_n = a/n^\alpha$ with $\alpha \in (1/2, 1)$. Instead of taking (X_n^α) as an estimate of ϑ, for which $n^{\alpha/2}(X_{n+1}^\alpha - \vartheta)$ is asymptotically $N(0, a\sigma^2/(2A))$-distributed, Polyak and Ruppert considered the averaged process $\overline{X}_n^\alpha := \frac{1}{n}\sum_{i=1}^n X_i^\alpha$. According to Polyak and Juditsky [11], $\sqrt{n}(\overline{X}_{n+1}^\alpha - \vartheta)$ is asymptotically $N(0, \sigma^2/A^2)$-distributed for any $a > 0$ and $A > 0$. It is to be expected that an Edgeworth expansion of the more complicated process $\sqrt{n}(\overline{X}_{n+1}^\alpha - \vartheta)$ will be valid under the mild assumptions $a > 0$ and $A > 0$ as well.

(b) *The influence of the starting random variable* X_1. It is worth to note that, according to [5], X_1 contributes a term of order $O(n^{-aA})$ to the asymptotic expansion of X_n. Hence, if $aA \leq 1$, the expansion in Theorem 3 cannot be valid.

(c) *The moment assumption*. In view of results in [2] and [12] it seems reasonable that the moment assumptions imposed on the V_n's in Theorems 3 and 4 can be weakened.

(d) *Different normalization*. Instead of $\sigma_n = \sqrt{\mathrm{var}(X_{n+1})}$ one may choose another sequence of normalizing factors τ_n of $X_{n+1} - \vartheta$. If we assume $\tau_n/\sigma_n = r_1 + r_2/n + o(n^{-1})$ with fixed real numbers r_1 and r_2, expansion (1) turns out to be

$$P\left(\frac{X_{n+1}-\vartheta}{\tau_n} \leq x\right) = \Phi(r_1x) + \frac{1}{\sqrt{n}} p_1(r_1x)\phi(r_1x) + \frac{1}{n}\left(p_2(r_1x) + r_2\right)\phi(r_1x) + o\left(\frac{1}{n}\right)$$

uniformly in $x \in \mathbb{R}$. As an example consider $\tau_n = 1/\sqrt{n}$. One can show that

$$\sigma_n = \frac{1}{\sqrt{n}} M_2 + \frac{1}{n^{3/2}}(M_{2,2} - M_1^2) + o(n^{-3/2})$$

with $M_2 = a\sigma/\sqrt{2aA - 1}$ and the numbers M_1 and $M_{2,2}$ as given in [6]. Then $r_1 = (2aA - 1)/(a\sigma)$ and $r_2 = -(M_{2,2} - M_1^2)/(2M_2^{3/2})$.

Assume a standard normal random variable N. The α-quantiles z_α and x_α related to the distributions of N and $|N|$ are given by $P(N \le z_\alpha) = \alpha$ and $P(|N| \le x_\alpha) = \alpha$, respectively ($\alpha \in (0, 1)$). Now Theorems 3 and 4 can be used to derive expansions for the coverage probabilities of one- and two-sided confidence intervals for ϑ, which are defined by

$$I_{1,n}(\alpha) := (X_{n+1} - \sigma_n z_\alpha,\ \infty),$$
$$I_{2,n}(\alpha) := (X_{n+1} - \sigma_n x_\alpha,\ X_{n+1} + \sigma_n x_\alpha).$$

Since $p_1(x)$ and $\phi(x)$ are *even* functions, the term related to $1/\sqrt{n}$ disappears in the expansion concerning the symmetric interval $I_{2,n}(\alpha)$. Observe that σ_n is often unknown in practice and must be estimated. In this case the following result has to be refined.

Corollary 6. *Under the assumptions of Theorem* 3 *it holds uniformly in* $\alpha \in (0, 1)$

$$P(\vartheta \in I_{1,n}(\alpha)) = \alpha - \frac{1}{\sqrt{n}}\left(k_{1,2} + \frac{1}{6}k_{3,1}(z_\alpha^2 - 1)\right)\phi(z_\alpha) + o\left(\frac{1}{\sqrt{n}}\right),$$
$$P(\vartheta \in I_{2,n}(\alpha)) = \alpha + o\left(\frac{1}{\sqrt{n}}\right).$$

Corollary 7. *Under the assumptions of Theorem* 4 *it holds uniformly in* $\alpha \in (0, 1)$

$$P(\vartheta \in I_{1,n}(\alpha)) = \alpha + \frac{1}{\sqrt{n}} p_1(z_\alpha)\phi(z_\alpha) + \frac{1}{n} p_2(z_\alpha)\phi(z_\alpha) + o\left(\frac{1}{n}\right),$$
$$P(\vartheta \in I_{2,n}(\alpha)) = \alpha + \frac{2}{n} p_2(x_\alpha)\phi(x_\alpha) + o\left(\frac{1}{n}\right).$$

Consider some asymptotically standard normal statistic $(X_{n+1} - \vartheta)/\sigma_n$ and its quantile function $Q_n \colon [0, 1] \to \overline{\mathbb{R}}$ defined by

$$Q_n(\alpha) = \inf\left\{x \in \mathbb{R} \colon P\left(\frac{X_{n+1} - \vartheta}{\sigma_n} \le x\right) \ge \alpha\right\}.$$

Following Cornish and Fisher this quantile function can be expanded formally in powers of $1/\sqrt{n}$ by

$$Q_n(\alpha) = \Phi^{-1}(\alpha) + \frac{1}{\sqrt{n}}\, q_1(\Phi^{-1}(\alpha)) + \frac{1}{n}\, q_2(\Phi^{-1}(\alpha)) + \ldots$$

Applying the Edgeworth expansion of $(X_{n+1} - \vartheta)/\sigma_n$ one can show formally [9] that

$$q_1(x) = -p_1(x) = k_{1,2} + \frac{1}{6}k_{3,1}(x^2 - 1)$$

with polynomial p_1 and constants $k_{1,2}$, $k_{3,1}$ as introduced above.

Proposition 8. *Assume that the* (*arbitrary*) *process* (X_n) *permits the Edgeworth expansion*

$$P\left(\frac{X_{n+1} - \vartheta}{\sigma_n} \le x\right) = \Phi(x) + \frac{1}{\sqrt{n}}\, p_1(x)\phi(x) + \delta_n \tag{2}$$

uniformly in $x \in \mathbb{R}$ *with* $\delta_n = o(1/\sqrt{n})$. *Then for all* $\varepsilon \in (0, \frac{1}{2})$ *it holds*

$$Q_n(\alpha) = \Phi^{-1}(\alpha) - \frac{1}{\sqrt{n}}\, q_1(\Phi^{-1}(\alpha)) + O\left(|\delta_n| + \frac{1}{n}\right) \tag{3}$$

uniformly in $\alpha \in (\varepsilon, 1 - \varepsilon)$.

Remark 9. **(a)** Under the assumptions of Theorems 3 or 4 the remainder term in (3) is of order $o(1/\sqrt{n})$ or $O(1/n)$, respectively.

(b) The expansion (3) cannot hold uniformly on the interval $[0,1]$. If, for example, $q_1(\Phi^{-1}(\alpha)) \to \infty$ for $\alpha \to \infty$, then the left-hand side of (3) would converge to ∞, the right-hand side however would converge to $-\infty$.

An Edgeworth expansion can be used to correct the distribution function of an asymptotically standard normal statistic $(X_{n+1} - \vartheta)/\sigma_n$ in order to obtain higher rates of convergence in the central limit theorem. The following theorem can be proved similarly to the last theorem by using Taylor expansions in the Edgeworth expansion.

Proposition 10. *Assume that an (arbitrary) process (X_n) possesses the Edgeworth expansion*

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \le x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \delta_n$$

uniformly in $x \in \mathbb{R}$. Then for any bounded interval $I \subset \mathbb{R}$

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \le x - \frac{1}{\sqrt{n}} p_1(x)\right) = \Phi(x) + O\left(|\delta_n| + \frac{1}{n}\right) \tag{4}$$

uniformly in $x \in I$.

Remark 11. **(a)** Remark 9(a) applies here as well.

(b) Since the parameters occurring in σ_n and p_1 are usually unknown, statement (4) with parameters σ_n and p_1 replaced by appropriate estimators is desirable. This problem was treated by Hall [8] for a statistic of type $f\left(\frac{1}{n}\sum_{i=1}^n V_i\right)$ with a sufficiently smooth function f.

A stochastic expansion of a statistic in the sense of [1] or [3] is a bivariate polynomial in $1/\sqrt{n}$ and a standard normal random variable N.

Proposition 12. **(a)** *For an (arbitrary) process (X_n) assume the validity of the Edgeworth expansion*

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \le x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x) + o\left(\frac{1}{\sqrt{n}}\right) \tag{5}$$

uniformly in $x \in \mathbb{R}$. Then the distribution function of $(X_{n+1} - \vartheta)/\sigma_n$ can be approximated by the distribution function of

$$N + \frac{1}{\sqrt{n}}\left(c_1 N^2 + c_2\right) \tag{6}$$

up to an error of order $o(1/\sqrt{n})$ uniformly in $x \in \mathbb{R}$.

(b) *For an (arbitrary) process (X_n) assume the validity of the Edgeworth expansion*

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n} \le x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x) + \frac{1}{n} p_2(x) + o\left(\frac{1}{n}\right) \tag{7}$$

uniformly in $x \in \mathbb{R}$. Then the distribution function of $(X_{n+1} - \vartheta)/\sigma_n$ can be approximated by the distribution function of

$$N + \frac{1}{\sqrt{n}}\left(c_1 N^2 + c_2\right) + \frac{1}{n} c_3 N^3 \tag{8}$$

up to an error of order $o(1/n)$ uniformly in $x \in \mathbb{R}$.

In both cases the coefficients c_1, c_2, c_3 are given by $c_1 = k_{3,1}/6$, $c_2 = k_{1,2} - k_{3,1}/6$ and $c_3 = k_{4,1}/24 - k_{3,1}^2/18$.

Theorems 3 and 4 give conditions which imply the validity of assumptions (5) and (7), respectively.

3. PROOFS

Instead of computing and proving the validity of Edgeworth expansions for the R−M process itself, we approximate the R−M process by sums of multilinear forms in the observation errors, which are simpler to treat. The validity of Edgeworth expansions for these approximands is proved in [7].

To derive central limit theorems for stochastic approximation procedures Walk [17] rewrote the recursion in a non-recursive representation consisting of a weighted sum defined in terms of the observation errors and a remainder term of negligible size. To obtain higher order approximations of the distribution function, the R−M process will be approximated more accurately involving additional quadratic and cubic terms in the observation errors. These additional terms reflect the non-linear behavior of the regression function. Furthermore, the weights in the linear form have to be refined due to the non-linearity of the recursion.

Thereto we assume that the R−M process $(X_{n+1} - \vartheta)$ can be approximated reasonable well by sums

$$S_n := L_n^\circ + Q_n \qquad \text{or} \qquad T_n := L_n + Q_n + C_n$$

of linear, quadratic (and cubic) forms in the observations errors $V_1, \dots, V_n$ as given in [7]. More precisely, we require the following two conditions, which hold under the assumptions of Theorems 3 and 4, respectively, as ensured by Theorem 1 in [5] (there with $p = 1$).

Condition 13 (2nd order representation). *The R−M process can be represented by*

$$X_{n+1} - \vartheta = L_n^\circ + Q_n + \Delta_n$$

with

$$\underset{\varepsilon>0}{\exists} \quad \underset{\varepsilon'>0}{\forall} \quad P\Big(|\sqrt{n}\Delta_n| \geq \varepsilon' n^{-\frac{1}{2}-\varepsilon}\Big) = o\Big(\frac{1}{\sqrt{n}}\Big).$$

Condition 14 (3rd order representation). *The R−M process can be represented by*

$$X_{n+1} - \vartheta = L_n + Q_n + C_n + \Delta_n$$

with

$$\underset{\varepsilon>0}{\exists} \quad \underset{\varepsilon'>0}{\forall} \quad P\Big(|\sqrt{n}\Delta_n| \geq \varepsilon' n^{-1-\varepsilon}\Big) = o\Big(\frac{1}{n}\Big).$$

It is of central importance that (S_n) and (T_n) have valid Edgeworth expansions as it is stated in Theorems 2 and 3 of [7].

Proof of Theorem 4. Remember that $\sigma_n = \sigma(X_{n+1})$. Due to Condition 14 we have for $\varepsilon > 0$ sufficiently small

$$P\bigg(\bigg|\frac{\Delta_n}{\sigma_n}\bigg| \geq n^{-1-\varepsilon}\bigg) = o\Big(\frac{1}{n}\Big).$$

Together with the so-called *delta method* [9] this relation can be used to conclude that

$$\begin{aligned} P\bigg(\frac{X_{n+1}-\vartheta}{\sigma_n} \leq x\bigg) &= P\bigg(\frac{T_n}{\sigma_n} + \frac{\Delta_n}{\sigma_n} \leq x\bigg) \\ &\leq P\bigg(\frac{T_n}{\sigma_n} \leq x + n^{-1-\varepsilon}\bigg) + P\bigg(\bigg|\frac{\Delta_n}{\sigma_n}\bigg| > n^{-1-\varepsilon}\bigg) = P\bigg(\frac{T_n}{\sigma_n} \leq x\bigg) + o\Big(\frac{1}{n}\Big) \end{aligned}$$

and, because of

$$\begin{aligned} P\bigg(\frac{T_n}{\sigma_n} \leq x - n^{-1-\varepsilon}\bigg) &\leq P\bigg(\frac{T_n}{\sigma_n} \leq x - \frac{\Delta_n}{\sigma_n} \text{ and } \bigg|\frac{\Delta_n}{\sigma_n}\bigg| \leq n^{-1-\varepsilon}\bigg) + P\bigg(\bigg|\frac{\Delta_n}{\sigma_n}\bigg| > n^{-1-\varepsilon}\bigg) \\ &\leq P\bigg(\frac{T_n}{\sigma_n} + \frac{\Delta_n}{\sigma_n} \leq x\bigg) + P\bigg(\bigg|\frac{\Delta_n}{\sigma_n}\bigg| > n^{-1-\varepsilon}\bigg), \end{aligned}$$

that

$$P\left(\frac{X_{n+1}-\vartheta}{\sigma_n}\le x\right)\ge P\left(\frac{T_n}{\sigma_n}\le x-n^{-1-\varepsilon}\right)-P\left(\left|\frac{\Delta_n}{\sigma_n}\right|>n^{-1-\varepsilon}\right)$$
$$=P\left(\frac{T_n}{\sigma(T_n)}\le x\right)+o\left(\frac{1}{n}\right).$$

To check that

$$P\left(\frac{T_n}{\sigma_n}\le x\pm n^{-1-\varepsilon}\right)=P\left(\frac{T_n}{\sigma(T_n)}\le x\right)+o(1/n)$$

one can use $\sigma_n-\sigma(T_n)=o(n^{-3/2})$, which can be concluded from Theorem 1 in [5], together with a Taylor expansion of the Edgeworth terms in Theorem 3 of [7]. Finally, another application of Theorem 3 in [7] proves the theorem. □

Proof of Theorem 3. This proof is achieved in the same manner but relying on Condition 13 and Theorem 2 of [7]. □

Proof of Proposition 8. Let $z_\alpha:=\Phi^{-1}(\alpha)$ be the α-quantile of the $N(0,1)$-distribution and assume that

$$q_{\alpha,n}=z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)$$

is the α-quantile of the distribution of $(X_n-\vartheta)/\sigma_n$. It is sufficient to show that $r_n(z_\alpha)=O\big(|\delta_n|+\frac{1}{n}\big)$. For that purpose we plug $q_{\alpha,n}$ in (2) and perform a Taylor expansion

$$\begin{aligned}\alpha&=\Phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)\right)\\&\quad+\frac{1}{\sqrt{n}}p_1\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)\right)\phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)\right)+\delta_n\\&=\Phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)\right)+r_n(z_\alpha)\phi\big(\xi(n,z_\alpha)\big)\\&\quad+\frac{1}{\sqrt{n}}\left\{p_1\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)\right)+r_n(z_\alpha)p_1'\big(\zeta(n,z_\alpha)\big)\right\}\phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)\right)+\delta_n.\end{aligned}\tag{9}$$

For $l_n(z_\alpha):=\phi(\xi(n,z_\alpha))$ it holds

$$\inf\left\{l_n(z_\alpha)\colon\alpha\in(\varepsilon,1-\varepsilon),\,n\in\mathbb{N}\right\}=:l>0.\tag{10}$$

Furthermore we have $\sup\big\{|p_1'(\zeta(n,z_\alpha))|\colon\alpha\in(\varepsilon,1-\varepsilon),\,n\in\mathbb{N}\big\}<\infty$. Carrying out another Taylor expansion we can continue equation (9) with

$$\begin{aligned}&=\Phi(z_\alpha)-\frac{1}{\sqrt{n}}p_1(z_\alpha)\phi(z_\alpha)+\frac{1}{n}p_1^2(z_\alpha)\phi'(\eta(n,z_\alpha))+r_n(z_\alpha)l_n(z_\alpha)\\&\quad+\frac{1}{\sqrt{n}}\left\{p_1(z_\alpha)+\frac{1}{\sqrt{n}}p_1(z_\alpha)p_1'(\rho(n,z_\alpha))+r_n(z_\alpha)O(1)\right\}\phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)+r_n(z_\alpha)\right)+\delta_n\end{aligned}$$

and, observing $\sup\{|p_1^2(z_\alpha)\phi'(\eta(n,z_\alpha))|\colon\alpha\in(\varepsilon,1-\varepsilon),\,n\in\mathbb{N}\}<\infty$ and $\sup\{|p_1(z_\alpha)\ p_1'(\rho(n,z_\alpha))|\colon$ $\alpha\in(\varepsilon,1-\varepsilon),\,n\in\mathbb{N}\}<\infty$,

$$\begin{aligned}&=\Phi(z_\alpha)-\frac{1}{\sqrt{n}}p_1(z_\alpha)\phi(z_\alpha)+O\left(\frac{1}{n}\right)+r_n(z_\alpha)l_n(z_\alpha)\\&\quad+\frac{1}{\sqrt{n}}\left\{p_1(z_\alpha)\left(\phi\left(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)\right)+r_n(z_\alpha)\phi'\big(\tau(n,z_\alpha)\big)\right)+O\left(\frac{1}{\sqrt{n}}\right)+r_n(z_\alpha)O(1)\right\}\\&\quad+\delta_n\end{aligned}$$

and, since $\sup\left\{|\phi'(\tau(n,z_\alpha))|\colon \alpha\in(\varepsilon,1-\varepsilon),\ n\in\mathbb{N}\right\}<\infty$ and $\phi\big(z_\alpha-\frac{1}{\sqrt{n}}p_1(z_\alpha)\big)=\phi(z_\alpha)+O\big(\frac{1}{\sqrt{n}}\big)$, we finally arrive at

$$=\alpha+O\left(\frac{1}{n}\right)+r_n(z_\alpha)\left(l_n(z_\alpha)+O\left(\frac{1}{\sqrt{n}}\right)\right)+\delta_n$$

or

$$r_n(z_\alpha)=-\frac{O(\frac{1}{n})+\delta_n}{l_n(z_\alpha)+O(\frac{1}{\sqrt{n}})}=O\left(|\delta_n|+\frac{1}{n}\right)$$

because of (10). □

Proof of Proposition 12. It is sufficient to compute Edgeworth expansions of (6) and (8) and then to compare the respective expansions with (5) or (7). □

REFERENCES

1. O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics* (Chapman and Hall, London, 1989).
2. R. N. Bhattacharya and J. K. Ghosh, "On the Validity of the Formal Edgeworth Expansion", Ann. Statist., 434–451 (1978).
3. D. R. Cox and N. Reid, "Approximations to Noncentral Distributions", Canad. J. Statist. **15**, 105–114 (1987).
4. J. Dippon, *Asymptotische Entwicklungen des Robbins–Monro Prozesses*, Habilitation Thesis (Universität Stuttgart, 1998).
5. J. Dippon, "Higher Order Representations of the Robbins–Monro Process", J. Multiv. Anal. **90**, 301–326 (2004).
6. J. Dippon, "Moments and Cumulants in Stochastic Approximation" (in press).
7. J. Dippon, "Edgeworth Expansions for Stochastic Approximation Theory", Math. Methods Statist. **17**, 44–65 (2008).
8. P. Hall, "Inverting an Edgeworth Expansion", Ann. Statist. **11**, 569–576 (1983).
9. P. Hall, *The Bootstrap and Edgeworth Expansion* (Springer, New York, 1992).
10. B. T. Polyak, "New Method of Stochastic Approximation Type", Automat. and Remote Control **51**, 937–946 (1990).
11. B. T. Polyak and A. B. Juditsky, "Acceleration of Stochastic Approximation by Averaging", SIAM J. Control and Optimization **30**, 838–855 (1992).
12. J. Renz, *Konvergenzgeschwindigkeit und asymptotische Konfidenzintervalle in der stochastischen Approximation*, PhD thesis (Universität Stuttgart, 1991).
13. H. Robbins and S. Monro, "A Stochastic Approximation Method", Ann. Math. Statist. **22**, 400–407 (1951).
14. D. Ruppert, "Stochastic Approximation", in *Handbook of Sequential Analysis*, Ed. by G. K. Ghosh and P. K. Sen (Marcel Dekker, 1991), pp. 503–529.
15. J. Sacks, "Asymptotic Distribution of Stochastic Approximation Procedures", Ann. Math. Statist. **29**, 373–405 (1958).
16. J. H. Venter, "An Extension of the Robbins–Monro Procedure", Ann. Math. Statist., 181–190 (1967).
17. H. Walk, "An Invariance Principle for the Robbins–Monro Process in a Hilbert Space", Z. Wahrscheinlichkeitsth. verw. Gebiete **39**, 135–150 (1977).