J. Appl. Prob. **55,** 559–570 (2018)
*doi:*10.1017/jpr.2018.35
© *Applied Probability Trust* 2018

ASYMPTOTIC BEHAVIOR FOR THE ROBBINS–MONRO PROCESS

YU MIAO * ** AND
MANRU DONG,* *** *Henan Normal University*

Abstract

In this paper we study the Robbins–Monro procedure $X_{n+1} = X_n - an^{-1}Y_n$ with some fixed number $a > 0$ and establish the moderate deviation principle of the process $\{X_n\}$.

Keywords: Robbins–Monro process; moderate deviation; exponential inequality

2010 Mathematics Subject Classification: Primary 62F12
Secondary 60F10

1. Introduction

We study the stochastic approximation method of Robbins and Monro [13]. For any real x, suppose that we can perform an experiment with a response y_x having distribution F_x. The expected response is then

$$M(x) = \int_{-\infty}^{\infty} y \, \mathrm{d}F_x(y).$$

In many applications, for example to process control or bioassay, a real number α is chosen and the aim is to estimate an unknown θ satisfying

$$M(\theta) = \alpha.$$

The Robbins–Monro procedure is used to estimate θ. Let X_1 be an arbitrary initial estimate of θ which is updated by the recursion

$$X_{n+1} = X_n - a_n Y_n.$$

Here $\{a_n\}$ is a suitable positive sequence of real numbers and Y_n has distribution F_{X_n}. If $a_n = n^{-1}$, $\{Y_n\}$ are bounded random variables, and $M(x)$ satisfies one of the following two conditions:

- for some $\delta > 0$, $M(x) \le \alpha - \delta$ for $x < \theta$ and $M(x) \ge \alpha + \delta$ for $x > \theta$;
- $M(\theta) = \alpha$, $M'(\theta) > 0$, and $M(x)$ is nondecreasing.

Robbins and Monro [13] proved that X_n converges in the second mean to θ, i.e. $\mathbb{E}|X_n - \theta|^2 \to 0$, which implies that X_n converges to θ in probability.

Received 26 January 2017; revision received 2 December 2017.
* Postal address: College of Mathematics and Information Science, Henan Normal University, 46# East of Construction Road, Xinxiang, Henan, 453007, China.
** Email address: yumiao728@126.com
*** Email address: 15936501139@163.com

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

Assume that the function $M(x)$ satisfies the following conditions:

(a) $|M(x)| \le c + d|x|,\ c, d \ge 0$;

(b) $\sigma_x^2 = \int_{-\infty}^{\infty} (y - M(x))^2 \,\mathrm{d}F_x(y) < \infty$ for every $x \in \mathbb{R}$;

(c) $M(x) < \alpha$ for $x < \theta$ and $M(x) > \alpha$ for $x > \theta$;

(d) $\inf_{\delta_1 \le |x-\theta| \le \delta_2} |M(x) - \alpha| > 0$ for every pair of numbers $(\delta_1, \delta_2) \subset (0, \infty)$,

and if the sequence of real numbers $\{a_n\}$ satisfies

$$\sum_{n=1}^{\infty} a_n = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} a_n^2 < \infty,$$

then $X_n \to \theta$ almost surely (a.s.); see Blum [2].

Chung [4] showed that if $a_n = An^{-\gamma}$ for $A > 0$ and $\frac{1}{2} < \gamma < 1$, or $A > (2M'(\theta))^{-1}$ and $\gamma = 1$, then

$$n^{\gamma/2}(X_{n+1} - \theta) \xrightarrow{\mathrm{D}} N(0, \sigma^2(\gamma, A)),$$

where

$$\sigma^2(\gamma, A) = \begin{cases} \dfrac{A\sigma_\theta^2}{2M'(\theta)}, & \frac{1}{2} < \gamma < 1, \\ \dfrac{A^2\sigma_\theta^2}{2AM'(\theta) - 1}, & \gamma = 1, \end{cases}$$

σ_θ^2 is defined as in (b) above, and '$\xrightarrow{\mathrm{D}}$' denotes convergence in distribution. Hodges and Lehmann [9] improved some of Chung's results. Derman [6] used Chung's methods to obtain results for the Kiefer–Wolfowitz procedure and Burkholder [3] extended Chung's methods to obtain further results on the asymptotic distribution of the Kiefer–Wolfowitz procedure.

For other limit behaviors of the Robbins–Monro procedure, Gapoškin and Krasulina [8] studied the law of the iterated logarithm for the procedure. Woodroofe [16] derived results concerning the rate of approximation to the normal distribution of $\sqrt{n}(X_n - \theta)$, and also established bounds for the large deviation probabilities $\mathbb{P}(|X_n - \theta| > \varepsilon)$ for ε positive and n large.

In this paper we are interested in the following regression problem. In order to estimate the zero θ of an unknown regression function $M\colon \mathbb{R} \to \mathbb{R}$ whose function value $M(X_n)$ at X_n can be observed only with some error V_n by $Y_n = M(X_n) + V_n$, we use the Robbins–Monro procedure

$$X_{n+1} = X_n - \frac{a}{n} Y_n \tag{1.1}$$

with some fixed number $a > 0$. Sacks [14] proved the asymptotic normality of $\sqrt{n}(X_n - \theta)$ using the method of characteristic functions. Following this, Walk [15] showed the asymptotic normality of the Robbins–Monro process in a Hilbert space by representing X_n as a weighted sum of the observation errors and a remainder term of order $o(n^{-1/2})$ with respect to convergence in probability. This representation can be considered as a weak first-order approximation of the Robbins–Monro process. A strong first-order approximation with respect to almost sure convergence was derived by Kersting [11]. Dippon [7] obtained (weak) higher-order representations which involved sums of linear, quadratic, and cubic forms of the observation errors. Dahmani and Ait Saidi [5] studied the consistency of the Robbins–Monro algorithm within a mixing framework. Arab and Dahmani [1] presented the complete convergence of the Robbins–Monro algorithm under quasi-associated random noise.

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

Our main aim in this paper is to establish the moderate deviations principle for the Robbins–Monro algorithm. To the best of the authors' knowledge, this is the first study on the large and moderate deviations for the Robbins–Monro algorithm. Moderate deviation estimations, as well as large deviation estimations, arise from the requirements of statistics. The large deviation offers a precise estimation associated to the law of large numbers, and the moderate deviation is often used to provide further estimations related to the central limit theorem and the law of the iterated logarithm. In order to provide a more explicit explanation, we assume that $\{X_n\}$ is a sequence of independent and identically distributed random variables. Set $\mu = \mathbb{E}X_1$ and $\bar{X}_n = n^{-1}\sum_{k=1}^{n} X_k$. For many purposes in statistics, one needs to study the limit behavior of

$$\mathbb{P}(\bar{X}_n - \mu \in A_n), \tag{1.2}$$

where A_n is a Borel subset of $\mathbb{R}$ and represents the deviation of $\bar{X}_n$ from μ. If $A_n = A$ for all n, (1.2) is the well-known large deviation. If $A_n = n^{-1/2}A$, the central limit theorem yields the limit behavior of (1.2). If $A_n = (b_n/\sqrt{n})A$, where b_n satisfies $b_n \to \infty$ and $b_n/\sqrt{n} \to 0$, (1.2) becomes the so-called moderate deviation. Hence, the moderate deviation is an intermediate estimation between the central limit theorem and the large deviation. In the following section we first state some useful assumptions and notation, and then establish a key exponential inequality of the Robbins–Monro algorithm. Finally, we present the moderate deviation principle.

2. Main results

2.1. Assumptions and notation

Assumption 2.1. *The error sequence $\{V_n\}$ is uniformly bounded and independent and identically distributed, i.e. there exists a positive constant L such that $|V_n| \le L$ a.s. and $\mathbb{E}V_n = 0$ for all n. Furthermore, let X_1 be independent in $\{V_n\}$ and $|X_1| \le L$ a.s.*

Assumption 2.2. *The function $M(\cdot)$ is a Borel measurable function with $M(\theta) = \alpha$ and, for all $x \ne \theta$,*

$$(x - \theta)(M(x) - \alpha) > 0. \tag{2.1}$$

Assumption 2.3. *For some positive constants K_1 and K_2, and for all x,*

$$K_1|x - \theta| \le |M(x) - \alpha| \le K_2|x - \theta|. \tag{2.2}$$

Assumption 2.4. *Let $\alpha_1 = M'(\theta)$ and assume that there exists a positive constant M_0 such that, for all $x \in \mathbb{R}$,*

$$|M''(x)| < M_0.$$

Assumption 2.5. *Let b_n be a sequence of real numbers satisfying the following conditions:*

$$b_n \to \infty \quad \text{and} \quad b_n = O(n^{1/2(1+\gamma)}) \quad \text{for some } \gamma > 0.$$

Remark 2.1. In [4], Assumptions 2.2–2.4 are the sufficient conditions needed for the asymptotic normal of the sequence $\{X_n\}$ to hold.

Let $a > 0$ and denote

$$\beta_{mn}(a) := \begin{cases} \displaystyle\prod_{j=m+1}^{n} (1 - aj^{-1}), & 0 \le m < n, \\ 1, & m = n. \end{cases}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

Lemma 2.1. *The array* $\{\beta_{mn}(a),\ 1 \le m \le n,\ n \ge 1\}$ *has the following elementary properties.*

(i) *For any* $n \ge 1,\ (2a-1) \vee 1 \le m \le n$, *we have*

$$C(a)\exp\Big(-\sum_{j=m+1}^{n}\frac{a}{j}\Big) \le \beta_{mn}(a) \le \exp\Big(-\sum_{j=m+1}^{n}\frac{a}{j}\Big), \tag{2.3}$$

where

$$C(a) := \exp\Big(-\frac{a^2}{(2a-1)\vee 1}\Big).$$

(ii) *For* $a > \frac{1}{2}$, *let*

$$t_n = \Big(\sum_{m=1}^{n} m^{-2}\beta_{mn}^2(a)\Big)^{-1/2}.$$

Then we have

$$\lim_{n\to\infty}\frac{t_n^2}{(2a-1)n} = 1. \tag{2.4}$$

Proof. (i) From the inequalities

$$1-x \le \mathrm{e}^{-x} \quad \text{and} \quad 1-x \ge \mathrm{e}^{-x-x^2} \quad \text{for all } 0 < x < \tfrac{1}{2},$$

we have

$$\exp\Big(-\sum_{j=m+1}^{n}\frac{a}{j}-\sum_{j=m+1}^{n}\frac{a^2}{j^2}\Big) \le \beta_{mn}(a) \le \exp\Big(-\sum_{j=m+1}^{n}\frac{a}{j}\Big).$$

The left-hand side of (2.3) can be obtained by using the elementary inequality

$$\sum_{j=m+1}^{n}\frac{a^2}{j^2} \le \frac{a^2}{(2a-1)\vee 1}.$$

(ii) For $k > 2$, using the inequalities

$$\ln(n-1) - \ln k \le \sum_{i=k}^{n}\frac{1}{i} \le \ln n - \ln(k-1), \tag{2.5}$$

we have, for any $n \ge 1$, $(2a-1)\vee 1 \le m \le n$,

$$C(a)\Big(\frac{m}{n}\Big)^a \le \beta_{mn}(a) \le \Big(\frac{m+1}{n-1}\Big)^a. \tag{2.6}$$

From (2.6), it is easy to check that, for any $\varepsilon > 0$, there exists a positive constant N_0 such that

$$1-\varepsilon \le (2a-1)n\sum_{m=N_0}^{n} m^{-2}\beta_{mn}^2(a) \le 1+\varepsilon. \tag{2.7}$$

Furthermore, using (2.6) again, for any $(2a-1)\vee 1 \le m \le N_0$, we have

$$n\beta_{mn}^2(a) \to 0 \quad \text{as } n \to \infty. \tag{2.8}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

For the $1 \le m \le (2a-1) \vee 1$ case, noting that

$$n\beta_{mn}^2(a) = \prod_{k=m}^{(2a-1)\vee 1}\left(1-\frac{a}{k}\right)\prod_{k=(2a-1)\vee 1+1}^{n}\left(1-\frac{a}{k}\right),$$

we see that the limit (2.8) also holds. Hence, from (2.7) and (2.8), claim (2.4) can be proved. This completes the proof. □

Remark 2.2. From the limit (2.8), it is easy to see that, for $a > \frac{1}{2}$ and for any $m \ge 0$, we have

$$t_n\beta_{mn}(a) \to 0 \quad \text{as } n \to \infty. \tag{2.9}$$

Throughout the following sections, we can assume that $\alpha = \theta = 0$ without loss of generality and let C be a positive constant which might not be the same in each of its appearances.

2.2. Exponential inequality

In this subsection we state an exponential inequality for the process $\{X_n\}$, which plays an important role in order to prove the moderate deviation principle of the sequence $\{X_n\}$.

Proposition 2.1. *Suppose that Assumptions 2.1–2.3 hold. Then, for any $\varepsilon > 0$, there exists a constant $\delta > 0$ such that, for all large enough n, we have*

$$\mathbb{P}(|X_n| > \varepsilon) \le 2\exp\left(-\frac{C\varepsilon^2\delta n}{1-\delta}\right)\left(1+\left(1-\exp\left(-\frac{C\varepsilon^2}{1-\delta}\right)\right)^{-1}\right). \tag{2.10}$$

Here δ can be chosen to satisfy

$$\delta < \exp\left(-\frac{2(T+\varepsilon)}{aK_1\varepsilon}\right),$$

where $T = T(a, K_1, L)$ is a positive constant.

Remark 2.3. From the proof of Proposition 2.1 (immediately following Lemma 2.4), if the assumption of the identical distribution in Assumption 2.1 is removed then (2.10) still holds.

Lemma 2.2. *For any $\varepsilon > 0$ and $T > 0$, there exists a positive constant δ such that, for all large enough n,*

$$\mathbb{P}(|X_n| > \varepsilon) \le \mathbb{P}\left(a\left|\sum_{k=[\delta n]}^{n-1}\frac{V_k}{k}\right| \ge 2\varepsilon\right) + \sum_{k=[\delta n]}^{n-1}\mathbb{P}\left(a\left|\sum_{i=k+1}^{n-1}\frac{V_i}{i}\right| \ge \frac{\varepsilon}{4}\right) + \mathbb{P}(|X_{[\delta n]}| > T),$$

where $[\delta n]$ denotes the integral part of δn.

Proof. We state the proof only for the form $\mathbb{P}(X_n > \varepsilon)$ since the other form $\mathbb{P}(X_n < -\varepsilon)$ is similar. For any $\varepsilon > 0$, $T > 0$, we can choose

$$\delta = \delta(\varepsilon, T, L) < \exp\left(-\frac{2(T+\varepsilon)}{aK_1\varepsilon}\right)$$

so small such that, for all large enough n,

$$T - a\sum_{i=[\delta n]}^{n-1}\frac{K_1\varepsilon/2}{i} < -\varepsilon, \tag{2.11}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

and define the events

$$B(\varepsilon) = \left\{X_{[\delta n]} \geq \tfrac{1}{2}\varepsilon, \ldots, X_{n-1} \geq \tfrac{1}{2}\varepsilon, X_n \geq \varepsilon\right\},$$
$$A_0^{(n)}(\varepsilon) = A_0(\varepsilon) = \left\{X_1 \geq \tfrac{1}{2}\varepsilon, \ldots, X_{n-1} \geq \tfrac{1}{2}\varepsilon, X_n \geq \varepsilon\right\},$$
$$A_k^{(n)}(\varepsilon) = A_k(\varepsilon) = \left\{X_k < \tfrac{1}{2}\varepsilon, X_{k+1} \geq \tfrac{1}{2}\varepsilon, \ldots, X_{n-1} \geq \tfrac{1}{2}\varepsilon, X_n \geq \varepsilon\right\},$$

where $k = 1, 2, \ldots, n-1$. Hence, it is easy to see that

$$\mathbb{P}(X_n > \varepsilon) \leq \mathbb{P}(X_{[\delta n]} \leq T, B(\varepsilon)) + \mathbb{P}(X_n > \varepsilon, (B(\varepsilon))^c) + \mathbb{P}(X_{[\delta n]} > T).$$

From (2.1) and (2.2), we see that

$$K_1 x \leq M(x) \leq K_2 x \quad \text{if } x > 0, \qquad K_2 x \leq M(x) \leq K_1 x \quad \text{if } x < 0.$$

Hence, by (2.11), we have

$$\begin{aligned}
\mathbb{P}(X_{[\delta n]} \leq T, B(\varepsilon)) &= \mathbb{P}\left(\varepsilon \leq X_{[\delta n]} - \sum_{k=[\delta n]}^{n-1} \frac{aY_k}{k}, T > X_{[\delta n]} \geq \frac{\varepsilon}{2}, \ldots, X_{n-1} \geq \frac{\varepsilon}{2}\right) \\
&\leq \mathbb{P}\left(\varepsilon \leq T - a\sum_{k=[\delta n]}^{n-1} \frac{K_1\varepsilon/2 + V_k}{k}\right) \\
&\leq \mathbb{P}\left(a\sum_{k=[\delta n]}^{n-1} \frac{V_k}{k} \leq T - \varepsilon - a\sum_{k=[\delta n]}^{n-1} \frac{K_1\varepsilon/2}{k}\right) \\
&\leq \mathbb{P}\left(a\sum_{k=[\delta n]}^{n-1} \frac{V_k}{k} \leq -2\varepsilon\right).
\end{aligned}$$

Next, we note that

$$X_{k+1} \leq X_k - \frac{aM(X_k)}{k} + \frac{aL}{k} \leq X_k + \frac{aL}{k} + \max\left(0, -\frac{aK_2X_k}{k}\right).$$

Then, for any $\varepsilon > 0$, there exists a positive constant K_0 such that, for all $k \geq K_0$, we have

$$\left\{X_k \leq \tfrac{1}{2}\varepsilon\right\} \subset \left\{X_{k+1} \leq \tfrac{3}{4}\varepsilon\right\}.$$

Hence, for all large enough n, we have

$$\begin{aligned}
&\mathbb{P}(X_n > \varepsilon, (B(\varepsilon))^c) \\
&\quad\leq \sum_{k=[\delta n]}^{n-1} \mathbb{P}(A_k(\varepsilon)) \\
&\quad= \sum_{k=[\delta n]}^{n-1} \mathbb{P}\left(X_k < \frac{\varepsilon}{2}, X_{k+1} \geq \frac{\varepsilon}{2}, \ldots, X_{n-1} \geq \frac{\varepsilon}{2}, X_n \geq \varepsilon\right) \\
&\quad= \sum_{k=[\delta n]}^{n-1} \mathbb{P}\left(X_k < \frac{\varepsilon}{2}, X_{k+1} \geq \frac{\varepsilon}{2}, \ldots, X_{n-1} \geq \frac{\varepsilon}{2}, X_{k+1} - a\sum_{i=k+1}^{n-1} \frac{M(X_i) + V_i}{i} \geq \varepsilon\right)
\end{aligned}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

$$\le \sum_{k=[\delta n]}^{n-1} \mathbb{P}\Big(a \sum_{i=k+1}^{n-1} \frac{V_i}{i} \le X_{k+1} - a \sum_{i=k+1}^{n-1} \frac{K_1\varepsilon/2}{i} - \varepsilon, X_k < \frac{\varepsilon}{2}\Big)$$
$$\le \sum_{k=[\delta n]}^{n-1} \mathbb{P}\Big(a \sum_{i=k+1}^{n-1} \frac{V_i}{i} \le -\frac{\varepsilon}{4}\Big).$$

From the above discussions, Lemma 2.2 can be obtained. □

Lemma 2.3. *The recursion X_n is bounded, i.e. there exists a positive constant C such that*

$$|X_n| \le C(n^{-aK_1}|X_1| + 1).$$

Proof. From (2.6), we have

$$|X_n| = \Big|X_1 \prod_{k=1}^{n-1}\Big(1 - \frac{aM(X_k)}{kX_k}\Big) + \sum_{i=1}^{n-1} \frac{aV_i}{i} \prod_{k=i+1}^{n-1}\Big(1 - \frac{aM(X_k)}{kX_k}\Big)\Big|$$
$$\le |X_1| \prod_{k=1}^{n-1}\Big|1 - \frac{aK_1}{k}\Big| + \sum_{i=1}^{n-1} \frac{aL}{i} \prod_{k=i+1}^{n-1}\Big|1 - \frac{aK_1}{k}\Big|$$
$$\le C|X_1|n^{-aK_1} + Cn^{-aK_1} \sum_{i=1}^{n-1} \frac{(i+1)^{aK_1}}{i}$$
$$\le C(n^{-aK_1}|X_1| + 1).$$ □

Lemma 2.4. (Hoeffding's inequality; see [10, Theorem 2].) *If $X_1, X_2, \dots, X_n$ are independent and $a_i \le X_i \le b_i$ $(i = 1, 2, \dots, n)$, then, for $t > 0$,*

$$\mathbb{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^{n}(X_k - \mathbb{E}X_k)\Big| \ge t\Big) \le 2\exp\Big(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big).$$

Proof of Proposition 2.1. From Lemma 2.3, we can choose $T = T(a, K_1, L) > 0$ so that $|X_n| \le T$ for all n. By using the inequalities in (2.5), for any $\varepsilon > 0$, we can choose $\delta = \delta(\varepsilon, T, L)$ so small such that, for all large enough n,

$$T - a \sum_{i=[\delta n]}^{n-1} \frac{K_1\varepsilon/2}{i} < -\varepsilon.$$

Then, from Lemmas 2.2–2.4, for all large enough n, we have

$$\mathbb{P}(|X_n| > \varepsilon) \le \mathbb{P}\Big(a\Big|\sum_{k=[\delta n]}^{n-1} \frac{V_k}{k}\Big| \ge 2\varepsilon\Big) + \sum_{k=[\delta n]}^{n-1} \mathbb{P}\Big(a\Big|\sum_{i=k+1}^{n-1} \frac{V_i}{i}\Big| \ge \frac{\varepsilon}{4}\Big)$$
$$\le 2\exp\Big(-\frac{2\varepsilon^2}{a^2L^2\sum_{k=[\delta n]}^{n-1} 1/k^2}\Big) + 2\sum_{k=[\delta n]}^{n-1} \exp\Big(-\frac{\varepsilon^2}{32a^2L^2\sum_{i=k+1}^{n-1} 1/i^2}\Big)$$
$$\le 2\exp\Big(-\frac{C\varepsilon^2 n}{\delta^{-1} - 1}\Big) + 2\sum_{k=[\delta n]}^{n-1} \exp\Big(-\frac{C\varepsilon^2}{k^{-1} - n^{-1}}\Big)$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

$$\le 2\exp\left(-\frac{C\varepsilon^2 n}{\delta^{-1}-1}\right)+2\sum_{k=[\delta n]}^{n-1}\exp\left(-\frac{C\varepsilon^2 k}{1-\delta}\right)$$
$$\le 2\exp\left(-\frac{C\varepsilon^2\delta n}{1-\delta}\right)+\frac{2\exp(-C\varepsilon^2\delta n/(1-\delta))}{1-\exp(-C\varepsilon^2/(1-\delta))}$$
$$\le 2\exp\left(-\frac{C\varepsilon^2\delta n}{1-\delta}\right)\left(1+\left(1-\exp\left(-\frac{C\varepsilon^2}{1-\delta}\right)\right)^{-1}\right). \qquad \square$$

2.3. Moderate deviation principle

In this subsection we state the moderate deviation principle of the recursion $\{X_n\}$.

Theorem 2.1. *Under Assumptions 2.1–2.5, let $a\alpha_1 > 1$. Then, for any $r > 0$, we have*

$$\lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}(h_n|X_{n+1}| \ge rb_n) = -\frac{r^2}{2\sigma^2},$$

where $\sigma^2 = \mathrm{var}(V_1)$ and $h_n = (a^2\sum_{m=1}^n m^{-2}\beta_{mn}^2(a\alpha_1))^{-1/2}$.

Remark 2.4. In general, the exponential integrable condition of the errors $\{V_n\}$ is enough to guarantee that the moderate deviation principle of some partial sums holds. However, one key point is to prove that the term $I_{2,n}$ (see (2.13)) can be neglected in the sense of the moderate deviation principle, and to this end, we need to use the exponential inequality (2.10) for $\{X_n\}$. The method to obtain (2.10) is based on the boundedness of the errors $\{V_n\}$, so the boundedness for the errors in Theorem 2.1 cannot be weakened to the exponential integrable condition.

Proof. From (1.1), we have

$$\begin{aligned} X_{n+1} &= X_n - \frac{a}{n}(M(X_n)+V_n) \\ &= X_n - \frac{a}{n}\left(M'(\theta)X_n + \frac{1}{2}M''(\eta_n)X_n^2 + V_n\right) \\ &= X_n\left(1-\frac{a\alpha_1}{n}\right) - \frac{a}{2n}M''(\eta_n)X_n^2 - \frac{a}{n}V_n, \end{aligned} \tag{2.12}$$

where $|\eta_n| \le |X_n|$ a.s. for every n. Iteration of (2.12) yields

$$\begin{aligned} X_{n+1} &= \prod_{k=1}^n\left(1-\frac{a\alpha_1}{k}\right)X_1 - \frac{a}{2}\sum_{k=1}^n k^{-1}\prod_{i=k+1}^n\left(1-\frac{a\alpha_1}{i}\right)M''(\eta_k)X_k^2 \\ &\quad - a\sum_{k=1}^n k^{-1}\prod_{i=k+1}^n\left(1-\frac{a\alpha_1}{i}\right)V_k \\ &=: I_{1,n} + I_{2,n} + I_{3,n}. \end{aligned} \tag{2.13}$$

First, from (2.9), we have

$$h_n|I_{1,n}| = h_n\beta_{0n}(a\alpha_1)|X_1| \le h_n\beta_{0n}(a\alpha_1)L \to 0.$$

Then it follows that, for any $r > 0$,

$$\lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}(h_n|I_{1,n}| > rb_n) = -\infty. \tag{2.14}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

Now we choose an array of positive real numbers $\{p_{kn},\ 1 \le k \le n,\ n = 1, 2, \ldots\}$ satisfying

$$\sum_{k=1}^{n} p_{kn} = 1. \tag{2.15}$$

For example, we can take

$$p_{kn} = d_n b_n^{1+\gamma} k^{a\alpha_1 - 2} n^{1/2 - a\alpha_1}, \qquad 1 \le k \le n,$$

where $\{d_n\}$ is a bounded real sequence, to guarantee that (2.15) holds. For any $r > 0$,

$$\begin{aligned}
\mathbb{P}(h_n |I_{2,n}| > r b_n) &\le \mathbb{P}\left(h_n M_0 \frac{a}{2} \sum_{k=1}^{n} k^{-1} \beta_{kn}(a\alpha_1) X_k^2 > r b_n \right) \\
&\le \sum_{k=1}^{n} \mathbb{P}\left(h_n M_0 \frac{a}{2} k^{-1} \beta_{kn}(a\alpha_1) X_k^2 > r b_n p_{kn} \right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X_k^2 > \frac{c_r k b_n p_{kn}}{\beta_{kn}(a\alpha_1) h_n} \right),
\end{aligned} \tag{2.16}$$

where $c_r := (M_0 a/2r)^{-1}$. On the one hand, it is not difficult to check that, for any fixed k,

$$\lim_{n\to\infty} \frac{c_r k b_n p_{kn}}{\beta_{kn}(a\alpha_1) h_n} = \lim_{n\to\infty} \frac{b_n^{2+\gamma}}{k} = \infty. \tag{2.17}$$

On the other hand, by Proposition 2.1, there exists a positive constant N_1 large enough such that

$$\begin{aligned}
\sum_{k=N_1}^{n} \mathbb{P}\left(X_k^2 > \frac{c_r k b_n p_{kn}}{\beta_{kn}(a\alpha_1) h_n} \right) &\le C \sum_{k=N_1}^{n} \exp\left(-Ck \frac{c_r k b_n p_{kn}}{\beta_{kn}(a\alpha_1) h_n} \right) \\
&\le Cn \exp(-C b_n^{2+\gamma}).
\end{aligned} \tag{2.18}$$

From (2.16)–(2.18), for any $r > 0$, we have

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}(h_n |I_{2,n}| > r b_n) = -\infty. \tag{2.19}$$

Next, we consider the term $I_{3,n}$. Letting

$$\omega_{kn} := a k^{-1} \beta_{kn}(a\alpha_1),$$

it is easy to see that (using (2.6))

$$\sum_{k=1}^{n} \omega_{kn}^2 \sim \frac{a^2}{(2a\alpha_1 - 1)n}, \qquad \sum_{k=1}^{n} \omega_{kn}^3 \sim \frac{a^3}{(3a\alpha_1 - 2)n^2}, \qquad \sum_{k=1}^{n} \omega_{kn}^4 \sim \frac{a^4}{(4a\alpha_1 - 3)n^3}.$$

Furthermore, from claims (2.14) and (2.19), it is necessary to prove that, for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}(h_n |I_{3,n}| > r b_n) = \lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(h_n \left| \sum_{k=1}^{n} \omega_{kn} V_k \right| > r b_n \right) = -\frac{r^2}{2\sigma^2}. \tag{2.20}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

In order to obtain claim (2.20), by the Gärtner–Ellis theorem, we need to prove that the following Crámer functional holds: for any $\lambda \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{E}\left(\exp\left(\lambda h_n b_n \sum_{k=1}^{n} \omega_{kn} V_k\right)\right) = \frac{\lambda^2 \sigma^2}{2}.$$

It is not difficult to see that

$$\begin{aligned}
&\left|\frac{1}{b_n^2} \log \mathbb{E}\left(\exp\left(\lambda h_n b_n \sum_{k=1}^{n} \omega_{kn} V_k\right)\right) - \frac{\lambda^2 h_n^2 E V_1^2}{2} \sum_{k=1}^{n} \omega_{kn}^2\right| \\
&\quad \le \sum_{k=1}^{n} \frac{1}{b_n^2} |\log \mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k)) - [\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k)) - 1]| \\
&\qquad + \sum_{k=1}^{n} \frac{1}{b_n^2} \left|[\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k)) - 1] - \frac{\lambda^2 h_n^2 b_n^2 \omega_{kn}^2 E V_1^2}{2}\right| \\
&\quad =: \Delta_1 + \Delta_2.
\end{aligned}$$

From Lemma 2.1, Assumption 2.1, and $a\alpha_1 > 1$, for all large enough n, there exists a positive constant C such that, for $1 \le k \le n$,

$$|\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k))| \le \mathbb{E}\left(\exp\left(\lambda h_n b_n \max_{1\le k\le n} |\omega_{kn}||V_k|\right)\right) \le C.$$

Hence, by the Taylor formula and Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\Delta_2 &= \sum_{k=1}^{n} \frac{1}{b_n^2} \left|[\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k)) - 1] - \frac{\lambda^2 h_n^2 b_n^2 \omega_{kn}^2 E V_1^2}{2}\right| \\
&\le \frac{1}{6b_n^2} \sum_{k=1}^{n} \mathbb{E}(|(\lambda h_n b_n \omega_{kn} V_k)^3 \exp(\lambda h_n b_n |\omega_{kn} V_k|)|) \\
&\le \frac{1}{6b_n^2} \sum_{k=1}^{n} (\lambda h_n b_n \omega_{kn})^3 (\mathbb{E}|V_k|^6)^{1/2} (\mathbb{E}(\exp(2\lambda h_n b_n |\omega_{kn} V_k|)))^{1/2} \\
&\le C b_n h_n^3 \sum_{k=1}^{n} \omega_{kn}^3 \\
&\to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Next, we consider the term Δ_1. First, again by the Taylor formula and Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\max_{1\le k\le n} |\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k)) - 1| \\
&\quad \le \max_{1\le k\le n} \mathbb{E}\left(\left|\frac{(\lambda h_n b_n \omega_{kn} V_k)^2}{2} \exp(\lambda h_n b_n \omega_{kn} V_k)\right|\right) \\
&\quad \le \max_{1\le k\le n} \frac{(\lambda h_n b_n \omega_{kn})^2}{2} (\mathbb{E} V_k^4)^{1/2} (\mathbb{E}(|\exp(2\lambda h_n b_n \omega_{kn} V_k)|))^{1/2}
\end{aligned}$$

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

$$\leq C \max_{1\leq k\leq n}\left(\frac{k^{a\alpha_1-1}b_n}{n^{a\alpha_1-1/2}}\right)^2$$
$$\to 0 \quad \text{as } n\to\infty.$$

So, for all large enough n, we have

$$\max_{1\leq k\leq n}|\mathbb{E}(\exp(\lambda h_n b_n \omega_{kn} V_k))-1|\leq \tfrac{1}{2}.$$

Using the elementary inequality

$$|\log(1+x)-x|\leq 2x^2, \qquad |x|\leq \tfrac{1}{2},$$

for all large enough n, we have

$$\begin{aligned}\Delta_1 &= \sum_{k=1}^{n}\frac{1}{b_n^2}|\log \mathbb{E}(\exp(\lambda h_n b_n\omega_{kn}V_k))-[\mathbb{E}(\exp(\lambda h_n b_n\omega_{kn}V_k))-1]|\\ &\leq \sum_{k=1}^{n}\frac{2}{b_n^2}[\mathbb{E}(\exp(\lambda h_n b_n\omega_{kn}V_k))-1]^2\\ &\leq \sum_{k=1}^{n}\frac{2}{b_n^2}\frac{(\lambda h_n b_n\omega_{kn})^4}{4}(\mathbb{E}V_k^4)(\mathbb{E}(|\exp(2\lambda h_n b_n\omega_{kn}V_k)|))\\ &\leq Cb_n^2h_n^4\sum_{k=1}^{n}\omega_{kn}^4\\ &\to 0 \quad \text{as } n\to\infty.\end{aligned}$$

From the limit (2.4), we have

$$\lim_{n\to\infty}\frac{\lambda^2h_n^2\mathbb{E}V_1^2}{2}\sum_{k=1}^{n}\omega_{kn}^2=\frac{\lambda^2\sigma^2}{2},$$

which, together with the above discussion, implies the desired results. □

References

[1] Arab, I. and Dahmani, A. (2016). Complete convergence of stochastic approximation algorithm in R^d under random noises. *Sequential Anal.* **35,** 216–225.

[2] Blum, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25,** 382–386.

[3] Burkholder, D. L. (1956). On a class of stochastic approximation processes. *Ann. Math. Statist.* **27,** 1044–1059.

[4] Chung, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25,** 463–483.

[5] Dahmani, A. and Ait Saidi, A. (2010). Consistency of Robbins Monro's algorithm within a mixing framework. *Bull. Soc. R. Sci. Liège* **79,** 131–140.

[6] Derman, C. (1956). An application of Chung's lemma to the Kiefer–Wolfowitz stochastic approximation procedure. *Ann. Math. Statist.* **27,** 532–536.

[7] Dippon, J. (2004). Higher order representations of the Robbins–Monro process. *J. Multivariate Anal.* **90,** 301–326.

[8] Gapoškin, V. F. and Krasulina, T. P. (1975). The law of the iterated logarithm in stochastic approximation processes. *Theory Prob. Appl.* **19,** 844–850.

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press

[9] Hodges, J. L., Jr. and Lehmann, E. L. (1956). Two approximations to the Robbins-Monro process. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955*, Vol. I, University of California Press, Berkeley, pp. 95–104.
[10] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58,** 13–30.
[11] Kersting, G. D. (1978). A weak convergence theorem with application to the Robbins–Monro process. *Ann. Prob.* **6,** 1015–1025.
[12] Major, P. and Révész, P. (1973). A limit theorem for the Robbins–Monro approximation. *Z. Wahrscheinlichkeitsth.* **27,** 79–86.
[13] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22,** 400–407.
[14] Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29,** 373–405.
[15] Walk, H. (1977). An invariance principle for the Robbins-Monro process in a Hilbert space. *Z. Wahrscheinlichkeitsth.* **39,** 135–150.
[16] Woodroofe, M. (1972). Normal approximation and large deviations for the Robbins–Monro process. *Z. Wahrscheinlichkeitsth.* **21,** 329–338.

https://doi.org/10.1017/jpr.2018.35 Published online by Cambridge University Press