The Annals of Probability
1978, Vol. 6, No. 6, 1015–1025

A WEAK CONVERGENCE THEOREM WITH APPLICATION TO THE ROBBINS-MONRO PROCESS

BY GÖTZ D. KERSTING
University of Göttingen

In this paper the asymptotic distribution of a sequence of random variables $(X_n)_{n\in\mathbb{N}}$, given by the recursion

$$X_{n+1} = X_n(1 - a_n{}^2 g(X_n)) + a_n Y_n,$$

is considered, where (Y_n) is a sequence of independent identically distributed random variables, $g : \mathbb{R} \to \mathbb{R}$ is a positive continuous function, and (a_n) is a sequence of positive numbers, going to zero. One application to the Robbins-Monro process is discussed, in which the function g will not be constant. Here the asymptotic distribution is no longer normal.

1. Introduction. In this paper a new method for calculating limiting distributions of stochastic processes is introduced. Let Y_n, $n = 1, 2, \ldots$ be a sequence of independent, identically distributed random variables with zero mean and finite variance. Define $X_n = n^{-\frac{1}{2}}\sum_{i=1}^{n} Y_i$. Then

$$X_{n+1} = X_n\big(1 - 1/2(n+1)^{-1} + 0(n^{-2})\big) + (n+1)^{-\frac{1}{2}} Y_{n+1}.$$

It is easily shown that the term $0(n^{-2})$ may be neglected if one is only interested in the limiting distribution of X_n. We shall more generally look at processes (X_n) which satisfy the recursion

$$X_{n+1} = X_n\big(1 - a_n^2 g(X_n)\big) + a_n Y_n$$

where g is a continuous positive function and (a_n) a sequence of positive numbers. We give a method for finding under some restrictions the limiting distribution of (X_n), which in general will no longer be normal. An application to the Robbins-Monro process is discussed in which the function g is not constant. (We discuss a case where the right and left derivatives of the regression function M exist, but are not necessarily equal, at the unique point θ where $M(\theta) = 0$; θ is to be estimated. See Section 3.)

We shall need no heavy machinery like characteristic functions. Thus our method may serve as a new easy way to prove the central limit theorem for sums of independent, identically distributed random variables with finite variance.

2. The limit theorem. We start with

LEMMA 2.1. *Let α_n, $\beta_n (n \geq 1)$ be nonnegative numbers such that $\alpha_n \to 0$, $\sum_{n=1}^{\infty}\alpha_n = \infty$ and for large n*

$$\beta_{n+1} \leq \beta_n(1 - c\alpha_n) + d\alpha_n$$

Received February 24, 1976; revised January 19, 1978.
AMS 1970 *subject classifications.* 60F05, 62L20.
Key words and phrases. Weak convergence, Robbins-Monro process.

with $c, d > 0$. Then

$$\limsup_{n\to\infty} \beta_n \leqslant d/c.$$

If $\beta_{n+1} \leqslant \beta_n(1 - c\alpha_n) + o(\alpha_n)$ then $\lim_{n\to\infty} \beta_n = 0$.

PROOF. Choose $\varepsilon > 0$. If $\beta_n \geqslant (d + \varepsilon)/c$, then

$$\beta_{n+1} \leqslant \beta_n - \frac{d+\varepsilon}{c} c\alpha_n + d\alpha_n \tag{2.1}$$
$$= \beta_n - \varepsilon\alpha_n$$

for n large enough. Thus, if $\beta_n \geqslant (d + \varepsilon)/c$ for all $n \geqslant n_0$, then $\beta_n \to -\infty$, since $\sum_{n=1}^{\infty}\alpha_n = \infty$. This is a contradiction. Thus there is an increasing sequence (n_k) of natural numbers containing just those numbers n with the property $\beta_n \leqslant (d + \varepsilon)/c$. For $n_k < n < n_{k+1}$ we get from (2.1)

$$\beta_n \leqslant \beta_{n-1} \leqslant \cdots \leqslant \beta_{n_k+1}$$
$$\leqslant \beta_{n_k} + d\alpha_{n_k} \leqslant (d + \varepsilon)/c + d\alpha_{n_k}.$$

Thus $\limsup_{n\to\infty} \beta_n \leqslant (d + \varepsilon)/c$. Letting $\varepsilon \to 0$ we get the desired result. The second statement follows immediately from the first. □

THEOREM 2.2. *Let $Y_1, Y_2, \ldots$ be a sequence of independent, identically distributed random variables with zero mean and finite variance σ^2. Let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable function with bounded derivative g' such that*

(i) $g'(x) \geqslant 0$ *for all* $x > 0$,
$g'(x) \leqslant 0$ *for all* $x < 0$;

(ii) $0 < d_1 \leqslant g(x) \leqslant d_2 < \infty$ *for all* $x \in \mathbb{R}$.

Let (a_n) be a sequence of nonnegative numbers such that

(iii) $\sum_{n=1}^{\infty} a_n^2 = \infty$

(iv) $\sum_{n=1}^{\infty} a_n^3 < \infty$.

If X_1 is a random variable with finite second moment and independent of Y_n, $n \geqslant 1$, and if (X_n) for $n \geqslant 1$ is given by

$$X_{n+1} = X_n\big(1 - a_n^2 g(X_n)\big) + a_n Y_n,$$

then X_n has a limiting distribution function F on $\mathbb{R}$ and its density (with respect to the Lebesgue measure) is

$$f(x) = C \exp(-h(x))$$

where $h(x) = 2\sigma^{-2}\int_0^x zg(z)\,dz$ and C is a normalizing constant.

(Note that from (ii) follows $h(x) \geqslant \sigma^{-2} d_1 x^2$. Thus all moments of F exist.)

PROOF. First suppose $g'(x) = 0$ for all x such that $|x| \geqslant D$ with a certain $D > 0$. We first show that there is an $N_1 \in \mathbb{N}$ such that for all $x, y \in \mathbb{R}$ and all $n \geqslant N_1$:

$$|x(1 - g(x)a_n^2) - y(1 - g(y)a_n^2) \leqslant |x - y|(1 - d_1 a_n^2). \tag{2.2}$$

Look at the function $s_n(x) = x(1 - g(x)a_n^2)$. Then

$$\begin{aligned} s_n'(x) &= 1 - g(x)a_n^2 - a_n^2 x g'(x) \\ &\geqslant 1 - d_2 a_n^2 - a_n^2 D \sup_x |g'(x)|. \end{aligned}$$

There is an N_1 such that $s_n'(x) > \frac{1}{2}$ and $(1 - g(x)a_n^2) > \frac{1}{2}$ for all $n \geqslant N_1$ and all $x \in \mathbb{R}$.

Now suppose $0 \leqslant x \leqslant y$. Because of condition (i) $g(x) \leqslant g(y)$. Thus for $n \geqslant N_1$

$$x(1 - g(x)a_n^2) \leqslant y(1 - g(y)a_n^2) \leqslant y(1 - g(x)a_n^2);$$

thus

$$\begin{aligned} &|x(1 - g(x)a_n^2) - y(1 - g(y)a_n^2)| \\ &\qquad \leqslant |x - y|(1 - g(x)a_n^2) \leqslant |x - y|(1 - d_1 a_n^2); \end{aligned}$$

thus (2.2) is true. Essentially the same argument holds if $x \leqslant y \leqslant 0$. If $x \leqslant 0 \leqslant y$, we have for $n \geqslant N_1$

$$\begin{aligned} x(1 - d_1 a_n^2) &\leqslant x(1 - g(x)a_n^2) \leqslant 0 \leqslant y(1 - g(y)a_n^2) \\ &\leqslant y(1 - d_1 a_n^2), \end{aligned}$$

which again leads to (2.2).

We now define a new process X_n', $n \geqslant N$, on the same probability space as X_n (by enlarging this space, if necessary), where $N \in \mathbb{N}$ is greater than N_1, such that X_N' has the distribution function F given in the theorem statement and is independent of Y_n, $n \geqslant N$, and such that

$$X_{n+1}' = X_n'(1 - g(X_n')a_n^2) + a_n Y_n$$

for $n \geqslant N$. By (2.2) we get for $n \geqslant N$

$$\begin{aligned} |X_{n+1} - X_{n+1}'| &= |X_n(1 - g(X_n)a_n^2) - X_n'(1 - g(X_n')a_n^2)| \\ &\leqslant |X_n - X_n'|(1 - d_1 a_n^2). \end{aligned}$$

Thus by condition (iii) and Lemma 2.1 $X_n - X_n' \to 0$ almost surely. Thus if the distribution of X_n' is near to F for large n, then the same will be true for X_n.

Now assume that Y_n takes on only finitely many values. Thus there are numbers r_i, $i = 1, \ldots, m$ such that for $p_i = P(Y_n = r_i)$

$$\Sigma_{i=1}^m p_i = 1, \qquad \Sigma_{i=1}^m p_i r_i = 0, \qquad \Sigma_{i=1}^m p_i r_i^2 = \sigma^2.$$

Take a fixed $n \geqslant N$ and choose $\varepsilon > 0$ such that

$$P(X_n' < x) \leqslant F(x) + \varepsilon \quad \text{for all } x \in R. \tag{2.3}$$

We shall show that there is a constant A (not depending on N or n) such that

(2.4) $$P(X'_{n+1} < x) \leqslant F(x) + \varepsilon + Aa_n^3.$$

By definition of X'_{n+1} we get

$$P(X'_{n+1} < x) = \Sigma_{i=1}^m p_i P\big(X'_n(1 - g(X'_n)a_n^2) < x - a_n r_i\big).$$

Since $s_n(x)$, as defined above, is continuous and strictly increasing for $n \geqslant N_1$, and since $s_n \to \pm\infty$ as $x \to \pm\infty$, there is exactly one α_i, $i = 1, \cdots, m$, such that

(2.5) $$\alpha_i(1 - g(\alpha_i)a_n^2) = x - a_n r_i.$$

Then by (2.3)

$$\begin{aligned} P(X'_{n+1} < x) &= \Sigma_{i=1}^m p_i P(X'_n < \alpha_i) \\ &\leqslant \Sigma_{i=1}^m p_i F(\alpha_i) + \varepsilon \\ &= F(x) + \varepsilon + \Sigma_{i=1}^m p_i \textstyle\int_x^{\alpha_i} f(y)\,dy. \end{aligned}$$

Now by a Taylor expansion

$$\begin{aligned} f(y) &= C\exp(-h(y)) \\ &= C\exp(-h(x)) - 2C\sigma^{-2}xg(x)\exp(-h(x))(y - x) \\ &\quad + f''(x + \delta_y(y - x))\frac{(y - x)^2}{2} \end{aligned}$$

with $0 < \delta_y < 1$. Thus

(2.6) $$\begin{aligned} \textstyle\int_x^{\alpha_i} f(y)\,dy &\leqslant C\exp(-h(x))(\alpha_i - x) \\ &\quad - C\sigma^{-2}xg(x)\exp(-h(x))(\alpha_i - x)^2 \\ &\quad + \sup_{|y-x|\leqslant|\alpha_i - x|}|f''(y)|\frac{|\alpha_i - x|^3}{6}. \end{aligned}$$

Now from (2.5),

$$\begin{aligned} \alpha_i &= (x - r_i a_n)(1 - g(\alpha_i)a_n^2)^{-1} \\ &= (x - r_i a_n)\big[1 + g(\alpha_i)a_n^2 + (g(\alpha_i))^2 a_n^4(1 - g(\alpha_i)a_n^2)^{-1}\big]. \end{aligned}$$

Thus there is an $A_1 > 0$ (note in the following that the constants A_V will not depend on N or n) such that

(2.7) $$|(\alpha_i - x) - (xg(\alpha_i)a_n^2 - r_i a_n)| \leqslant A_1(1 + |x|)a_n^3.$$

This implies

(2.8) $$|(\alpha_i - x)| \leqslant A_2(1 + |x|)a_n.$$

By the mean value theorem, since g' is bounded,

$$|g(\alpha_i) - g(x)| \leqslant A_3(1 + |x|)a_n.$$

Thus from (2.7)

(2.9) $$|(\alpha_i - x) - (xg(x)a_n^2 - r_i a_n)| \leqslant A_4(1 + x^2)a_n^3.$$

From (2.7) we see that

$$|(\alpha_i - x) + r_i a_n| \leqslant A_5(1 + |x|)a_n^2.$$

Using this and (2.8) we obtain

$$\begin{aligned} |(\alpha_i - x)^2 - r_i^2 a_n^2| &= |(\alpha_i - x) + r_i a_n| \cdot |(\alpha_i - x) - r_i a_n| \\ &\leqslant A_5(1 + |x|)a_n^2[A_2(1 + |x|)a_n + |r_i|a_n] \\ &\leqslant A_6(1 + x^2)a_n^3. \end{aligned}$$

From (2.6) we get by means of (2.8), (2.9) and the last inequality

$$\begin{aligned} \textstyle\int_x^{\alpha_i} f(y)\,dy &\leqslant C\exp(-h(x))\big(xg(x)a_n^2 - r_i a_n + A_4(1 + x^2)a_n^3\big) \\ &\quad - C\sigma^{-2}xg(x)\exp(-h(x))\big(r_i^2 a_n^2 - A_6(1 + x^2)a_n^3\big) \\ &\quad + \sup_{|y-x|\leqslant|\alpha_i - x|}|f''(y)|A_2^3(1 + |x|)^3 a_n^3/6. \end{aligned}$$

Now $\exp(-h(x))|x|^2$ is bounded and $\sup_{|y-x|\leqslant|\alpha_i-x|}|f''(y)|\,|x|^3$ is bounded (in x), thus

$$\begin{aligned} \textstyle\int_x^{\alpha_i} f(y)\,dy \leqslant{}& C\exp(-h(x))\big(xg(x)a_n^2 - r_i a_n\big) \\ & - C\sigma^{-2}xg(x)\exp(-h(x))r_i^2 a_n^2 \\ & + Aa_n^3. \end{aligned}$$

Thus

$$\begin{aligned} \textstyle\sum_{i=1}^m p_i \int_x^{\alpha_i} f(y)\,dy \leqslant{}& C\exp(-h(x))xg(x)a_n^2 \textstyle\sum_{i=1}^m p_i \\ & - C\exp(-h(x))a_n \textstyle\sum_{i=1}^m r_i p_i \\ & - C\sigma^{-2}xg(x)\exp(-h(x))a_n^2 \textstyle\sum_{i=1}^m r_i^2 p_i \\ & + Aa_n^3 = Aa_n^3. \end{aligned}$$

This proves (2.4). Since $P(X_N' < x) = F(x)$, we get from (2.3) and (2.4) by induction

$$\limsup_{n\to\infty} P(X_n' < x) \leqslant F(x) + A\textstyle\sum_{n=N}^{\infty} a_n^3.$$

Thus, since $X_n - X_n' \to 0$ a.s., for any $\delta > 0$

$$\limsup_{n\to\infty} P(X_n < x) \leqslant F(x + \delta) + A\textstyle\sum_{n=N}^{\infty} a_n^3.$$

Letting $N \to \infty$, $\delta \to 0$

$$\limsup_{n\to\infty} P(X_n < x) \leqslant F(x).$$

Similarly

$$\liminf_{n\to\infty} P(X_n < x) \geqslant F(x).$$

This proves the theorem under the restrictions in the proof.

We now proceed to the general case where Y_i is not longer discrete and $g'(x)$ need not vanish for large $|x|$. Choose $\eta > 0$. Construct identically distributed random variables Y_i' which are Y_i-measurable and thus independent, which assume only finitely many values, and which satisfy

$$E(Y_i') = 0, \qquad E((Y_i' - Y_i)^2) \leqslant \eta.$$

Further choose $\bar{g} : R \to R$ satisfying the conditions of the theorem and such that for some $D > 0$

$$\bar{g}'(x) = 0 \qquad \text{for all} \quad |x| \geqslant D$$
$$|g(x) - \bar{g}(x)| \leqslant \eta \qquad \text{for all} \quad x \in R.$$

Define $X_1' = X_1$,

$$X_{n+1}' = X_n'(1 - \bar{g}(X_n')a_n^2) + a_n Y_n'.$$

Then $E(X_{n+1}'^2) \leqslant E(X_n'^2)(1 - d_1 a_n^2) + \sigma'^2 a_n^2$, with $\sigma'^2 = E(Y_i'^2)$. Thus by Lemma 2.1

$$\limsup_{n\to\infty} E(X_n'^2) \leqslant \sigma'^2/d_1.$$

Similarly

$$\limsup_{n\to\infty} E(X_n^2) \leqslant \sigma^2/d_1.$$

Now by independence and (2.2), if n is large enough that $1 - \bar{g}(X_n)a_n^2 \geqslant 0$ and $1 - \bar{g}(X_n')a_n^2 \geqslant 0$,

$$\begin{aligned}
E((X_{n+1} - X_{n+1}')^2) &\\
&= E\left[(X_n(1 - \bar{g}(X_n)a_n^2) - X_n'(1 - \bar{g}(X_n')a_n^2))^2\right] \\
&\quad -2E\left[X_n a_n^2(g(X_n) - \bar{g}(X_n))\right]\left[X_n(1 - \bar{g}(X_n)a_n^2) - X_n'(1 - \bar{g}(X_n')a_n^2)\right] \\
&\quad +E\left[X_n^2 a_n^4(g(X_n) - \bar{g}(X_n))^2\right] + a_n^2 E((Y_n - Y_n')^2) \\
&\leqslant E((X_n - X_n')^2)(1 - d_1 a_n^2)^2 \\
&\quad +2a_n^2\eta(E(|X_n X_n'|) + E(X_n^2)) \\
&\quad +a_n^4\eta^2 E(X_n^2) + a_n^2\eta \\
&\leqslant E((X_n - X_n')^2)(1 - d_1 a_n^2) \\
&\quad +2a_n^2\eta\left(\frac{\sigma\sigma' + \sigma^2}{d_1}\right)(1 + o(1)) + \eta a_n^2.
\end{aligned}$$

Thus by Lemma 2.1

$$\limsup_{n\to\infty} E((X_n - X_n')^2) \leqslant \eta\left(2\frac{\sigma\sigma' + \sigma^2}{d_1} + 1\right)/d_1.$$

Since $\sigma' \to \sigma$, as $\eta \to 0$, this bound may be chosen arbitrarily small, say smaller than τ^3 for any $\tau > 0$.

By Tschebyscheff's inequality then

$$\limsup_{n\to\infty} P(|X_n - X_n'| \geqslant \tau) \leqslant \tau,$$

thus

$$\limsup_{n\to\infty} P(X_n < x) \leqslant \limsup_{n\to\infty} P(X_n' < x + \tau) + \tau$$
$$= C' \int_{-\infty}^{x+\tau} \exp\left(-2\int_0^y z\bar{g}(z)\, dz/\sigma'^2\right) dy + \tau.$$

We let $\eta \to 0$ and $\bar{g} \to g$. It follows that $\sigma' \to \sigma$ and $\tau \to 0$ so that

$$\limsup_{n\to\infty} P(X_n < x) \leqslant F(x).$$

Similarly,

$$\liminf_{n\to\infty} P(X_n < x) \geqslant F(x).$$

This proves the theorem. □

We shall see later that the theorem remains true for a larger class of functions $g : R \to R^+$ as described in the theorem by using a similar approximation argument as in the end of the proof.

3. Application to the Robbins-Monro process. The Robbins-Monro process, as introduced by Robbins and Monro (1951), deals with the problem of estimating the root θ of the equation

$$M(x) = 0$$

for an unknown measurable function M. The method is the following: choose an arbitrary X_1 and define a sequence of random variables X_n by

$$X_{n+1} = X_n - c_n Z_n,$$

where c_n is a sequence of nonnegative numbers and the random variable Z_n denotes a measurement of M at the point X_n so that

$$E(Z_n | X_1, \cdots, X_n) = M(X_n) \quad \text{a.s.} \tag{3.1}$$

Several authors proved convergence of X_n to θ for suitable choice of a_n. (See Blum (1954), Chung (1954); see Schmetterer (1961) for a more complete bibliography.)

If $c_n = n^{-1}$, Chung proved asymptotic normality of $n^{\frac{1}{2}}(X_n - \theta)$ under several assumptions; this was generalized later by several authors (Fabian (1968), Sacks (1958)). In all these papers one essential condition is that the derivative $M'(\theta)$ exists and

$$M'(\theta) > \tfrac{1}{2}. \tag{3.2}$$

In addition to other results Révész and Major (1973) got asymptotic normality of $(n/\log n)^{\frac{1}{2}}(X_n - \theta)$, if $M'(\theta) = \frac{1}{2}$ and a.s. convergence of $X_n^{M'(\theta)}$, if $0 < M'(\theta) < \frac{1}{2}$. In this paper we look at the case where $M'(\theta)$ does not exist, but the derivatives at θ from the right and left exist. (This problem was posed by Dvoretzky.)

We shall look only at the situation where the error of measurement $M(X_n) - Z_n$ is independent of X_n, i.e., we shall assume that $-M(X_n) + Z_n$, $n \geqslant 1$, is a sequence of independent, identically distributed random variables and that X_1 is independent from this sequence.

Now assume that

$$m_1 := \lim_{x\downarrow\theta} M(x)(x-\theta)^{-1},$$

and

$$m_2 := \lim_{x\uparrow\theta} M(x)(x-\theta)^{-1}$$

exist. Denote

$$Y_n = M(X_n) - Z_n.$$

We need the following lemma due to Chung (1954), page 466, and Venter (1966), page 1535.

LEMMA 3.1. *Let α_n be a sequence of real numbers such that for large n*

$$\alpha_{n+1} \leqslant \alpha_n(1 - cn^{-1}) + dn^{-1-\rho}$$

with $c, d, \rho > 0$. If $c > \rho$ then $\alpha_n = 0(n^{-\rho})$. If $c < \rho$ then $\alpha_n = 0(n^{-c})$.

THEOREM 3.2. *Let (X_n) be a Robbins-Monro process with $c_n = n^{-1}$ such that*

(i) $M(x)(x-\theta) > 0$ *for all* $x \neq \theta$,

(ii) $|M(x)| \leqslant A + B|x|$ *for all* $x \in \mathbb{R}$,

(iii) $\inf_{r\leqslant|x-\theta|\leqslant R}|M(x)| > 0$ *for all* $0 < r < R < \infty$,

(iv) $E(Y_n^2) = \sigma^2 < \infty, \quad E(X_1^2) < \infty$,

(v) $m_1, m_2 > \frac{1}{2}$.

Then $n^{\frac{1}{2}}(X_n - \theta)$ has a limiting distribution and its density is

$$f(x) = C\exp\left(-\frac{x^2(2m_1-1)}{2\sigma^2}\right) \quad \text{if} \quad x > 0,$$

$$f(x) = C\exp\left(-\frac{x^2(2m_2-1)}{2\sigma^2}\right) \quad \text{if} \quad x \leqslant 0$$

with

$$C = \left(\frac{2}{\pi\sigma^2}\right)^{\frac{1}{2}} \frac{(2m_1-1)^{\frac{1}{2}}(2m_2-1)^{\frac{1}{2}}}{(2m_1-1)^{\frac{1}{2}} + (2m_2-1)^{\frac{1}{2}}}.$$

PROOF. Suppose $\theta = 0$. Conditions (i)–(iv) imply almost sure convergence of X_n to 0. (This is Blum's theorem.) Define $m(x) = m_1$ for $x \geqslant 0$, $m(x) = m_2$ for $x < 0$. We assume that $M(x) - m(x)x$ is bounded for $x \in R$ and also that

$$|M(x)| \geqslant K_1|x|$$

with $K_1 > \frac{1}{2}$. If we prove the theorem under these conditions, then the general case follows by a trick of Hodges and Lehmann (see their paper and also the proof of Theorem 1′ in Sacks (1958)).

Now by independence

$$E(X_{n+1}^2) = E\left(X_n^2\left(1 - n^{-1}\frac{M(X_n)}{X_n}\right)^2\right) + \sigma^2 n^{-2}$$
$$\leqslant E(X_n^2)(1 - (2K_1 + o(1))n^{-1}) + \sigma^2 n^{-2}.$$

Since $2K_1 > 1$, by Lemma 3.1 with $\rho = 1$

$$E(X_n^2) = 0(n^{-1}).$$

Thus $E(|X_n|) = 0(n^{-\frac{1}{2}})$ and

$$P(|X_n| \geqslant \varepsilon) = 0(n^{-1})$$

for all $\varepsilon > 0$. Since $|M(x) - xm(x)| = o(x)$ and $\sup_x |M(x) - xm(x)| < \infty$ this implies

$$E(|M(X_n) - X_n m(X_n)|) = o(n^{-\frac{1}{2}}).$$

Now

$$X_{n+1} = X_n(1 - m(X_n)n^{-1}) + n^{-1}(X_n m(X_n) - M(X_n)) + n^{-1}Y_n.$$

Multiply this by $n^{\frac{1}{2}}$, and obtain

$$n^{\frac{1}{2}}X_{n+1} = (n-1)^{\frac{1}{2}}X_n\left(1 - \left(m(X_n) - \tfrac{1}{2}\right)n^{-1} + 0(n^{-2})\right)$$
$$+ n^{-\frac{1}{2}}(m(X_n)X_n - M(X_n)) + n^{-\frac{1}{2}}Y_n.$$

Define $\bar{X}_n$ by $\bar{X}_1 = 0$ and

$$\bar{X}_{n+1} = \bar{X}_n\left(1 - n^{-1}\left(m(\bar{X}_n) - \tfrac{1}{2}\right)\right) + n^{-\frac{1}{2}}Y_n.$$

As in the proof of (2.2), if n is large enough, then

$$|x\left[1 - \left(m(x) - \tfrac{1}{2}\right)n^{-1}\right] - y\left[1 - \left(m(y) - \tfrac{1}{2}\right)n^{-1}\right]|$$
$$\leqslant |x - y|(1 - \bar{m}n^{-1})$$

uniformly in x and y where $\bar{m} = \min(m_1, m_2) - \frac{1}{2} > 0$. Then, since $m(X_n) = m(X_n(n-1)^{\frac{1}{2}})$, for large enough n

$$|n^{\frac{1}{2}}X_{n+1} - \bar{X}_{n+1}| \leqslant |(n-1)^{\frac{1}{2}}X_n - \bar{X}_n|(1 - \bar{m}n^{-1})$$
$$+ (n-1)^{\frac{1}{2}}|X_n|0(n^{-2})$$
$$+ n^{-\frac{1}{2}}|X_n m(X_n) - M(X_n)|.$$

Thus

$$E\left(|n^{\frac{1}{2}}X_{n+1} - \bar{X}_{n+1}|\right) \leqslant E\left(|(n-1)^{\frac{1}{2}}X_n - \bar{X}_n|\right)(1 - \bar{m}n^{-1}) + o(n^{-1}).$$

By Lemma 2.1 $E(|(n-1)^{\frac{1}{2}}X_n - \overline{X}_n|) \to 0$. Thus it is sufficient to show that the limiting distribution of $\overline{X}_n$ is the one given in the theorem.

Now choose $\eta > 0$ and $g_\eta : R \to R$ such that

$$g_\eta(x) = m(x) - \tfrac{1}{2} \qquad \text{for} \quad |x| \geqslant \eta$$

such that g_η satisfies the conditions of Theorem 2.2, and such that $m(x) - \frac{1}{2} \geqslant g_\eta(x) \geqslant \overline{m} > 0$. Define the process $\overline{\overline{X}}_n$ by $\overline{\overline{X}}_1 = 0$ and

$$\overline{\overline{X}}_{n+1} = \overline{\overline{X}}_n\left(1 - n^{-1}g_\eta\left(\overline{\overline{X}}_n\right)\right) + n^{-\frac{1}{2}}Y_n.$$

Then as in Theorem 2.2

$$\begin{aligned}|\overline{X}_{n+1} - \overline{\overline{X}}_{n+1}| &\leqslant |\overline{X}_n\left(1 - g_\eta(\overline{X}_n)n^{-1}\right) - \overline{\overline{X}}_n\left(1 - g_\eta(\overline{X}_n)n^{-1}\right)| \\ &\quad + |\overline{X}_n|n^{-1}\left(m(\overline{X}_n) - \tfrac{1}{2} - g_\eta(\overline{X}_n)\right) \\ &\leqslant |\overline{X}_n - \overline{\overline{X}}_n|(1 - \overline{m}n^{-1}) + \eta n^{-1}|m_1 - m_2|.\end{aligned}$$

Thus by Lemma 2.1

$$\limsup_{n\to\infty} E\left(\overline{X}_n - \overline{\overline{X}}_n\right) \leqslant \eta\overline{m}^{b-1}|m_1 - m_2|.$$

Now from Theorem 2.2 with $a_n = n^{-\frac{1}{2}}$ we see that $\overline{\overline{X}}_n$ has a limiting distribution F_η and its density is

$$C_\eta \exp\left(-2\sigma^{-2}\textstyle\int_0^x y g_\eta(y)\,dy\right).$$

Thus by the argument at the end of the proof of Theorem 2.2, letting $\eta \to 0$, $g_\eta \to m - \frac{1}{2}$, $\overline{X}_n$ must converge to the distribution given in the theorem. The normalizing constant may easily be calculated. □

REFERENCES

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

[2] CHUNG, K. L. (1954). On stochastic approximation methods. *Ann. Math. Statist.* **25** 463–483.

[3] FABIAN, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327–1332.

[4] HODGES, J. L. and LEHMANN, E. L. (1956). Two approximations to the Robbins-Monro process. *Proc. Third Berkeley Symp. Math. Statist. Prob.* 95–104.

[5] KERSTING, G. (1978). On the rate of convergence of the Robbins-Monro process. Unpublished manuscript.

[6] MAJOR, P. and RÉVÉSZ, P. (1973). A limit theorem for the Robbins-Monro approximation. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** 79–86.

[7] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

[8] SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

[9] SCHMETTERER, L. (1961). Stochastic approximation. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 587–609.

[10] VENTER, J. H. (1966). On Dvoretzky stochastic approximation theorems. *Ann. Math. Statist.* **37** 1534–1544.

INSTITUT FÜR MATHEMATISCHE STATISTIK UND WIRTSCHAFTSMATHEMATIK
GÖTTINGEN UNIVERSITY
D-34 GÖTTINGEN
LOTZESTRASSEN 13
FEDERAL REPUBLIK OF GERMANY