Ann. Inst. Statist. Math.
34 (1982), Part A, 545-558

LOCATING THE MINIMUM OF A FUNCTION WHEN THE ERRORS OF OBSERVATION HAVE UNKNOWN DENSITY

THOMAS E. OBREMSKI

(Received June 30, 1981; revised Jan. 11, 1982)

Summary

In considering the problem of locating the point θ at which a function f achieves its minimum (or maximum) using the Kiefer-Wolfowitz (KW) stochastic approximation procedure, Abdelhamid [1] has shown that if the density g of the errors obtained in estimating functional values is known, then a transformation of observations leads to methods which under mild conditions have desirable asymptotic properties. We address the more general problem of locating the point of minimum of a function when g is unknown to the experimenter. In the procedure given in Theorem 4.1 we obtain the same asymptotic results as Abdelhamid in his version of the KW procedure.

1. Introduction

In many experimental situations, the experimenter is interested in estimating the point θ at which a function attains its minimum (or maximum). Often, the actual form of the function is unknown, but at each point x the experimenter is able to obtain an estimate Y_x of the functional value $f(x)$. Suppose that at each point x the error of observation, $Y_x - f(x)$, has density g. If the error random variable $Y_x - f(x)$ has mean zero and finite variance, the Kiefer-Wolfowitz (KW) [11] procedure may used to locate the point θ. For certain known densities, a procedure suggested by Abdelhamid [1] yielding smaller asymptotic second moment is applicable. It is for the case g unknown that we propose here a procedure, given in (2.6.3), which has the same asymptotic second moment as the Abdelhamid procedure.

More formally, consider the stochastic approximation procedure given by

$$(1.1) \qquad X_{n+1} = X_n - a_n c_n^{-1} Y_n , \qquad n = 1, 2, \cdots ,$$

where X_n, Y_n are random variables and a_n, c_n are positive numbers.

Included in (1.1) are both the KW procedure and the Robbins-Monro [13] procedure (RM).

Abdelhamid and, independently, Anbar [2] investigated the possible effect that transforming the observed random variables Y_n might have on the almost sure convergence and the asymptotic normality of $\{X_n\}$. Abdelhamid's investigation included both the KW and RM cases; Anbar's only the RM case.

Specifically they studied the asymptotic behavior of the procedure given by

$$X_{n+1}=X_n-a_n c_n^{-1}h(Y_n)\ , \qquad n=1, 2, \cdots , \tag{1.2}$$

where h was assumed to belong to a class C of Borel measurable functions which preserve both the almost sure convergence of X_n to θ and the asymptotic normality of $n^{\beta}(X_n-\theta)$, where in the RM case, θ is the unknown root of the function f and $\beta=1/2$, and in the KW case, θ is the point of minimum of f and β lies in the interval [1/4, 1/3], depending on the assumptions on f. Denote by F_n the sigma-algebra generated by $X_1, X_2, \cdots, X_n$. The following conditions were assumed for the random variables $V_n=Y_n-\mathrm{E}^{F_n}(Y_n)$: V_n are conditionally (given F_n) distributed according to a symmetric distribution function G admitting density g; g has derivative almost everywhere with respect to G; $0<I(g)=\int (g'(v)/g(v))^2 dG(v)<+\infty$.

Within the class C, they sought a function h^* which would minimize the second moment of the asymptotic distribution. It was known that in the case where g is normal, such an h was given by the identify function, that is, when g is known to be normal, (1.1) cannot be improved upon by transformation of observations. They found in general that within the class C, $h^*(v)=(-g'/g)(v)$, unique up to multiplicative constant. So for example if g is double exponential, then $(-g'/g)(v)= C\,\mathrm{sign}\,(v)$ with a constant $C>0$, and the optimal procedure is

$$X_{n+1}=X_n-a_n c_n^{-1}\,\mathrm{sign}\,(Y_n)\ , \qquad n=1, 2, \cdots , \tag{1.3}$$

first suggested by Fabian [5], [6].

Abdelhamid also suggested improvements in some cases where G is known but fails to satisfy all of the assumptions above.

Without assuming knowledge of the distribution G, Fabian [9] constructed an RM-type procedure which performs asymptotically as well as the transformed RM procedure (1.2) does when G is known. With $a_n=an^{-1}$, $c_n=1$ in the RM case and $cn^{-\gamma}$ in the KW case, a and c positive numbers, and γ in the interval [1/6, 1/4], Abdelhamid had derived values of a and c optimal in the sense of minimizing the second moment of the asymptotic distribution. In the RM case the optimal choice of

a is $(f'(\theta)I(g))^{-1}$. Fabian suggested methods of estimating $I(g)$, $-g'/g$, and $f'(\theta)$ and pointed out some of the problems inherent in such estimation.

The main purpose of this paper is to achieve asymptotic results in the KW case with G unknown which are as strong as those obtained by Abdelhamid. Much of the motivation for this present paper was provided by Fabian's 1973 paper which we shall refer to henceforth as I. In some places we were able to apply results obtained in I directly to the KW case. These places are indicated in the text. For example, much of the actual estimation of unknown parameters that is outlined in our main result, Theorem 4.1 is carried out as in I.

The same speeds of convergence Theorem 2.11 and asymptotic normality result (Theorem 3.1) as those obtained by Abdelhamid (Theorems 4.4 and 4.5) are achieved. The main result, Theorem 4.1, is a realization of the procedure suggested, indicating how to estimate the optimal values of a and c, as well as $I(g)$, $-g'/g$, $f''(\theta)$, and $f'''(\theta)$. We show in Sections 2 and 3 that our procedure has the same asymptotic properties as those given by Abdelhamid in the case when g is known.

In Abdelhamid's treatment of the KW case, and in many of the earlier treatments, the following two assumptions have appeared: First, there exist constants A and B such that

$$|f(x+1)-f(x)|<A|x-\theta|+B\,, \qquad \text{for every } x \text{ in } R\,, \tag{1.4}$$

and, secondly,

$$\mathrm{E}^{F_n}(V_n^2)\leqq\sigma^2\,, \qquad \text{for every natural number } n\,,$$

for a number σ and $V_n=Y_n-\mathrm{E}^{F_n}(Y_n)$. The latter assumption may be omitted here if in the truncation of the Y_n given in (2.6.4), y_n are chosen to be $(\log(n\vee 2))^{1-2\varepsilon_1}$. The use of truncation in (2.6.4) also enables us to weaken the assumption (1.4) to f being bounded on bounded intervals. Without the truncated term in the recursion relation (2.6.3) we would need to assume not only (1.4) but also a similar type of condition for $\mathrm{E}^{F_n}(-g'/g)(Y_n)$.

2. Almost sure convergence of the proposed stochastic approximation procedure

2.1. *Basic notation*

All random variables are assumed to be defined on a probability space $(\Omega, \mathcal{F}, P)$. Relations between random variables, including convergence, are meant to hold almost surely, unless specified otherwise.

The set of real numbers is denoted by R, positive reals by R^+, and the class of all Borel subsets of R by $\mathscr{B}$. The indicator function of a set S is denoted by $\mathscr{X}_S$. Let E denote expectation, and $\mathrm{E}^{\mathscr{F}}$ conditional expectation, given the σ-algebra $\mathscr{F}$. If $Z_1, \cdots, Z_n$ are random variables, then $\mathscr{F}(Z_1, \cdots, Z_n)$ denotes the σ-algebra induced by $Z_1, \cdots, Z_n$.

If $\{b_n\}$ is a sequence of numbers and $\{Z_n\}$ a sequence of random variables, then we write $Z_n = \mathcal{O}(b_n)$ if $\limsup |b_n^{-1} Z_n(\omega)| < +\infty$ for almost all ω. Similarly we write $Z_n = \mathcal{O}_u(b_n)$ if there exists a K in R and an integer n_0 with $|b_n^{-1} Z_n| \leqq K$, for all $n \geqq n_0$.

If φ is a function on R and a is in R^+, then for each x in R, $\varphi^a(x)$ denotes the difference $\varphi(x+a) - \varphi(x-a)$; if k is a natural number, $D^k\varphi(x)$ denotes the kth derivative of φ at x.

2.2. *Remark*

The following assumptions are listed for reference later. Assumptions 2.6 and 2.7 appear in the convergence results in this chapter, Assumption 2.8 in the asymptotic normality result, Theorem 3.1. Only Assumptions 2.3 and 2.4 appear in the main result, Theorem 4.1.

2.3. *Assumption*

Both θ and γ belong to R. We assume that f is a function on R such that either

$$D^2 f \text{ exists, is continuous in a neighborhood of } \theta, \text{ and } \gamma = 1/4\,,$$

or

$$D^3 f \text{ exists, is continuous in a neighborhood of } \theta, \text{ and } \gamma = 1/6\,.$$

We further assume that f is bounded on bounded intervals, that $D^2 f(\theta) = M > 0$, and that for every natural number k,

$$\sup_{-k < x-\theta < -1/k} \overline{D} f(x) < 0\,; \quad \inf_{1/k < x-\theta < k} \underline{D} f(x) > 0\,, \tag{2.3.1}$$

where $\overline{D}f(x)$ and $\underline{D}f(x)$ denote respectively the upper and lower derivatives of f at x.

2.4. *Assumption*

Assumption 2.3 holds. We assume that $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ are random variables, that $\mathscr{F}_n$ is a non-decreasing sequence of σ-algebras such that for each n, $\mathscr{F}_n$ contains the σ-algebra $\mathscr{F}(X_1, \cdots, X_n, Y_1, \cdots, Y_{n-1})$. For each n, C_n is a positive $\mathscr{F}_n$-measurable random variable with $c_n = C_n n^{-\gamma}$, and $f^{c_n}(X_n)$ is the $\mathscr{F}_n$-measurable random variable whose value at ω is

$$f(X_n(\omega) + c_n(\omega)) - f(X_n(\omega) - c_n(\omega))\,.$$

For each n, $Y_n - f^{c_n}(X_n)$ is conditionally, given $\mathcal{F}_n$, distributed according to a distribution function G which is symmetric, has zero expectation, has a density g which has a continuous derivative Dg everywhere on R. The density g is non-increasing on $[0, \infty)$ and $0 < I(g) = \int (g^{-1}D(g))^2 \cdot dG < +\infty$.

2.5. *Remark*

The assumption of symmetry of G is a natural one in a Kiefer-Wolfowitz type of procedure, where Y_n is an unbiased estimator of $f^{c_n}(X_n)$. This requirement is satisfied, for example, if the errors in estimating $f(X_n+c_n)$ and $f(X_n-c_n)$, respectively, are independent and identically distributed, given $\mathcal{F}_n$.

2.6. *Assumption*

Assumption 2.4 holds and h_n are measurable functions on $(\Omega \times R, \mathcal{F}_n \times \mathcal{B})$ such that for each ω, $h_n(\omega, \cdot)$ are odd, and are non-negative on $[0, +\infty)$. For each n, D_n is a non-negative $\mathcal{F}_n$-measurable random variable and

$$|h_n(\omega, t)| \leqq n^{\varepsilon_1} \chi_{(-n,n)}(t) , \quad (\log n)^{-\varepsilon_0} \leqq C_n \leqq (\log n)^{\varepsilon_0} , \quad D_n \leqq n^{\varepsilon_1} \tag{2.6.1}$$

with numbers ε_0, ε_1 satisfying $0 < \varepsilon_1 < \gamma/2$ and $0 < \varepsilon_0 < \varepsilon_1/2$. Note that for both possible values of γ, we can (and will) select a μ_γ such that

$$1/2 - \gamma - 2\varepsilon_1 > \mu_\gamma > 0 . \tag{2.6.2}$$

We shall write $h_n(t)$ for $h_n(\cdot, t)$, and $h_n(Y_n)$ for $h_n(\cdot, Y_n(\cdot))$.

The random variables $X_1, X_2, \cdots$ satisfy

$$X_{n+1} = X_n - (nc_n)^{-1}[D_n h_n(Y_n) + \log(n \vee 2)^{-1+\varepsilon_1} \tilde{Y}_n] \tag{2.6.3}$$

where

$$\tilde{Y}_n = (Y_n \vee (-y_n)) \wedge y_n \tag{2.6.4}$$

with $y_n = n^{\varepsilon_1}$ if G has finite second moment and $y_n = (\log(n \vee 2))^{1-2\varepsilon_1}$ otherwise.

2.7. *Assumption*

Assumption 2.6 holds. For almost all ω, $h_n(\omega, \cdot) \to -g^{-1}(Dg)$ on the set $\{t; g(t) > 0\}$ and

$$D_n \to (2MI(g))^{-1} . \tag{2.7.1}$$

2.8. *Assumption*

Assumption 2.7 holds and

$$\int [h_n(t+\eta_n(t))+g^{-1}(Dg)]^2 dG \to 0 \tag{2.8.1}$$

for every sequence $\{\eta_n\}$ of functions on $\Omega\times R$ with $|\eta_n|\leqq|f^{c_n}(X_n)|$ and such that, for almost all ω, $h_n(\omega, t+\eta_n(\omega, t))$ are Borel measurable with respect to t. The random variables $C_1, C_2, \cdots$ satisfy

$$C_n \to \begin{cases} \left[\left(\frac{3}{2}\right)(D^3f(\theta))^{-2}I^{-1}(g)\right]^{1/6} & \text{if } \gamma=1/6 \text{ and } D^3f(\theta)\neq 0 \\ C & \text{otherwise}, \end{cases} \tag{2.8.2}$$

where C is in R^+.

2.9. *Remark*

Suppose Assumption 2.6 holds. In the proof of Theorem 2.10 we shall require expressions for $\mathrm{E}^{\mathcal{F}_n}h_n(Y_n)$ and $\mathrm{E}^{\mathcal{F}_n}\tilde{Y}_n$. If k is a Borel measurable function, then the conditional expectation $\mathrm{E}^{\mathcal{F}_n}k(Y_n)$, provided it exists, is equal to $K(\mathrm{E}^{\mathcal{F}_n}Y_n)$ where $K(\Delta)=\int k(t+\Delta)g(t)dt$. For $k=h_n$ and $k(t)=(t\vee(-y_n))\wedge y_n$, several properties of K were established in I under the same conditions on g and h_n as we assumed in 2.6. So using the results (I3.1.1), (I3.1.2) and (I3.1.9) we have

$$\mathrm{E}^{\mathcal{F}_n}h_n(Y_n)=\Psi_n(f^{c_n}(X_n)), \qquad \mathrm{E}^{\mathcal{F}_n}\tilde{Y}_n=f^{c_n}(X_n)\kappa_n \tag{2.9.1}$$

where Ψ_n are functions satisfying

$$\Delta\Psi_n(\Delta)\geqq 0 \quad \text{for all } \Delta \text{ in } R, \tag{2.9.2}$$

$$\Delta^{-1}\Psi_n(\Delta)\leqq kn^{\epsilon_1} \text{ with a } k \text{ in } R^+, \quad \text{for all } \Delta\neq 0, \tag{2.9.3}$$

and κ_n (equal to $\phi_n'(\Delta_n)$ in I) are non-negative $\mathcal{F}_n$-measurable random variables with $\kappa_n\to 1$ on the set of all ω for which $\{f^{c_n}(X_n(\omega))\}$ is a bounded sequence.

2.10. *Theorem*

If Assumption 2.6 *holds, then* $(\log n)^\beta(X_n-\theta)\to 0$, *for every* $\beta>0$.

PROOF. Assume without loss of generality that $\theta=0$. Let $\varepsilon>0$. It is easy to see that there is a function φ on R such that $\varphi(x)=\varphi(-x)$ for all x, $\varphi=0$ on $[0, \varepsilon]$ and $\varphi>0$ on $(\varepsilon, +\infty)$, φ has a bounded second derivative and first derivative $\boldsymbol{D}$ satisfying $x\boldsymbol{D}(x)\geqq 0$ and $|\boldsymbol{D}(x)|\leqq|x|$ for all x, and $\boldsymbol{D}(x)=x$ for $x>2\varepsilon$.

We have $c_n=C_nn^{-\gamma}\leqq n^{\epsilon_1-\gamma}$ by (2.6.1) and it suffices to consider n so large that $c_n<\varepsilon/2$. Then by (2.9.1)

$$\boldsymbol{D}(X_n)\,\mathrm{E}^{\mathcal{F}_n}\tilde{Y}_n\geqq 0, \tag{2.10.1}$$

since $\kappa_n \geqq 0$ and sign $f^{c_n}(x) = \operatorname{sign} x$ for $|x| > c_n$ by (2.3.1). Define $B_n = (c_n^{-1} \cdot \boldsymbol{D}(X_n) \, \mathrm{E}^{\mathcal{F}_n} \tilde{Y}_n)^{1/2}$. Write (2.6.3) as $X_{n+1} = X_n - U_n$, and $N_n = \mathrm{E}^{\mathcal{F}_n} U_n$. Then with $\alpha_n = n^{-1}(\log n)^{-1+\varepsilon_1}$, we have $\boldsymbol{D}(X_n) N_n = A_n + \alpha_n B_n^2$, where by (2.9.2), $A_n \geqq 0$. So

$$\boldsymbol{D}(X_n) N_n \geqq \alpha_n B_n^2 \,. \tag{2.10.2}$$

Also, by (2.6.1), (2.6.2), and (2.6.4)

$$\mathrm{E}^{\mathcal{F}_n} U_n^2 = \mathcal{O}_u(n^{-1-2\mu_\gamma}) \quad \text{with} \quad \mu_\gamma > 0 \,. \tag{2.10.3}$$

Relations (2.10.2) and (2.10.3) show that conditions 2, 3, and 4 of Lemma 3.3, Fabian [8], are satisfied with $\gamma_n = \varepsilon_n = 0$, and $\beta_n = k(\log n)^{2\varepsilon_0}(n^{-1-2\mu_\gamma})$ with a k in R^+. Hence a subsequence $\{B_{n_i}\}$ of $\{B_n\}$ converges to 0 and the sequence $\{\varphi(X_n)\}$ converges to a random variable.

Let ω be a point at which both properties hold. Since $\varphi(X_{n_i}(\omega))$ converges, $X_{n_i}(\omega)$ is bounded, and then so is $f^{c_{n_i}}(X_{n_i}(\omega))$. Therefore $\kappa_{n_i}(\omega) \to 1$ and so $c_{n_i}^{-1}(\omega) \boldsymbol{D}(X_{n_i}(\omega)) f^{c_{n_i}}(X_{n_i}(\omega)) \to 0$. The latter convergence, the properties of $\boldsymbol{D}$, and (2.3.1) imply that $\limsup |X_{n_i}(\omega)| \leqq \varepsilon$. But, since $\varphi(X_n(\omega))$ converges, $\varphi(X_n(\omega)) \to 0$ and $\limsup |X_n(\omega)| \leqq \varepsilon$. The final relation holds for all ω in a set of probability one. Since ε was chosen arbitrary and positive, $X_n \to 0$. (Note that as a consequence, $\kappa_n \to 1$.)

Now suppose that $\mathcal{N}$ is a neighborhood of $x=0$ in which $D^2 f$ exists and is continuous if $\gamma = 1/4$, and in which $D^3 f$ exists and is continuous if $\gamma = 1/6$. Expanding $f^{c_n}(X_n)$ in powers of c_n in $\mathcal{N}$ we obtain, with a proper choice of ξ_n, that

$$c_n^{-1} f^{c_n}(X_n) = \alpha_n X_n + \xi_n c_n^{-1+1/2\gamma} \,, \tag{2.10.4}$$

where α_n and ξ_n are $\mathcal{F}_n$-measurable random variables, with

$$\alpha_n \to 2M, \quad \xi_n \to \xi_0 = \begin{cases} 0 & \text{if } \gamma = 1/4 \\ \frac{1}{3} f'''(0) & \text{if } \gamma = 1/6 \,. \end{cases} \tag{2.10.5}$$

Using expression (2.10.4) we obtain

$$\begin{aligned} N_n = {} & n^{-1}(\alpha_n X_n + \xi_n c_n^{-1+1/2\gamma}) \\ & \cdot [D_n(f^{c_n}(X_n))^{-1} \Psi_n(f^{c_n}(X_n)) + (\log n)^{-1+\varepsilon_1} \kappa_n] \,. \end{aligned} \tag{2.10.6}$$

For ε_0 as in (2.6.1), we obtain from (2.10.6) and (2.9.3) that

$$N_n = n^{-1}(\log n)^{-1+\varepsilon_0}(\alpha_n X_n + \xi_n c_n^{-1+1/2\gamma}) \delta_n \,, \tag{2.10.7}$$

with

$$0 \leqq \delta_n = \mathcal{O}_u((\log n)^{1-\varepsilon_0} n^{2\varepsilon_1}) \,, \qquad \delta_n \to +\infty \,. \tag{2.10.8}$$

Then $X_n - N_n = X_n(1 - n^{-1}(\log n)^{-1+\epsilon_0}\alpha_n\delta_n) - R_n$, where $R_n = n^{-1}(\log n)^{-1+\epsilon_0}\xi_n \cdot c_n^{-1+1/2\gamma}\delta_n$. Note that from (2.10.5), (2.10.8), (2.6.1) and (2.6.2) we have that

$$R_n = \mathcal{O}(n^{-1-\mu_\gamma}) , \qquad \mu_\gamma > 0 . \tag{2.10.9}$$

Now, eventually, depending on ω, $0 \leqq 1 - n^{-1}(\log n)^{-1+\epsilon_0}\alpha_n\delta_n \leqq 1$, $(1 - n^{-1}(\log n)^{-1+\epsilon_0}\alpha_n\delta_n)^2 \leqq 1 - n^{-1}(\log n)^{-1+\epsilon_0}\alpha_n\delta_n \leqq 1 - n^{-1}(\log n)^{-1+\epsilon_0}$, and

$$(X_n - N_n)^2 \leqq X_n^2(1 - n^{-1}(\log n)^{-1+\epsilon_0}) + 2|X_n R_n| + R_n^2 . \tag{2.10.10}$$

Writing $X_{n+1} = (X_n - N_n) - (U_n - N_n)$ we obtain from (2.10.10)

$$X_{n+1}^2 \leqq (1 - A_n)X_n^2 - 2V_n + W_n + T_n \tag{2.10.11}$$

with

$$A_n = n^{-1}(\log n)^{-1+\epsilon_0} , \tag{2.10.12}$$

$$V_n = (X_n - N_n)(U_n - N_n) , \qquad W_n = (U_n - N_n)^2 , \qquad T_n = 2|X_n R_n| + R_n^2 . \tag{2.10.13}$$

Suppose now β_n are positive numbers satisfying (eventually)

$$\beta_n^{-1}\beta_{n+1}(1 - A_n) \leqq 1 , \qquad \beta_{n+1}X_n = \mathcal{O}(n^{\mu_\gamma - \eta}) , \qquad \beta_n \leqq^{2\mu_\gamma - \eta} \quad \text{for an } \eta > 0 . \tag{2.10.14}$$

We shall show that under these conditions

$$\sum_{n=1}^{\infty} \beta_{n+1}W_n < +\infty , \quad \sum_{n=1}^{\infty} \beta_{n+1}T_n < +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \beta_{n+1}V_n < \infty . \tag{2.10.15}$$

This, (2.10.14), and (2.10.11) then easily imply

$$\beta_n X_n^2 = \mathcal{O}(1) . \tag{2.10.16}$$

The first relation in (2.10.15) follows from (2.10.3) since $\mathrm{E}\, W_n \leqq \mathrm{E}\, U_n^2$. The second relation follows since by (2.10.9), $R_n^2 = \mathcal{O}(n^{-2-2\mu_\gamma})$ and $\beta_{n+1}|X_n||R_n| = \mathcal{O}(n^{\mu_\gamma - \eta})\mathcal{O}(n^{-1-\mu_\gamma})$.

From (2.10.3), $\mathrm{E}^{\mathcal{F}_n} V_n^2 = (X_n - N_n)^2 \mathcal{O}_u(n^{-1-2\mu_\gamma})$. But $(X_n - N_n)^2 \leqq X_n^2 + T_n$ by (2.10.10) and $\sum_{n=1}^{\infty} \beta_{n+1}^2 T_n |\mathcal{O}_u(n^{-1-2\mu_\gamma})| \leqq \sum_{n=1}^{\infty} \beta_{n+1}T_n < +\infty$ as we have already shown. Concerning the other term, we have

$$\beta_{n+1}^2 X_n^2 \mathcal{O}_u(n^{-1-2\mu_\gamma}) = \mathcal{O}(n^{-1-\eta}) .$$

This shows that $\sum_{n=1}^{\infty} \beta_{n+1}^2\, \mathrm{E}^{\mathcal{F}_n} V_n^2 < +\infty$ and $\sum_{n=1}^{\infty} \beta_{n+1}V_n$ converges by the generalized Borel-Cantelli Lemma (Lemma 10, Dubins and Freedman [4]). Thus (2.10.15) holds.

Choose now $\beta_n=(\log n)^b$ so that (2.10.14) is satisfied and (2.10.16) holds. This proves the theorem.

2.11. *Theorem*

If Assumption 2.7 *holds, then* $n^{\beta}(X_n-\theta)\to 0$ *for every* $0<\beta<1/2-\gamma-2\varepsilon_1$.

PROOF. Relation (I3.1.16) can be rewritten as

$$\liminf \mu_n \geqq I(g) \tag{2.11.1}$$

with $\mu_n=f^{-1}(X_n)\Psi_n(f(X_n))$. This relation holds also in our case with $\mu_n=[f^{c_n}(X_n)]^{-1}\Psi_n[f^{c_n}(X_n)]$. So we obtain from (2.10.6) and (2.7.1) a strengthening of (2.10.7) to

$$N_n=n^{-1}k_n(\alpha_n X_n+\xi_n c_n^{-1+1/2\gamma}) \tag{2.11.2}$$

with $\liminf k_n \geqq (2M)^{-1}$, $k_n=\mathcal{O}_u(n^{2\varepsilon_1})$. Then, (2.10.11) holds with (2.10.12) strengthened to

$$A_n=2n^{-1}\alpha_n k_n' \tag{2.11.3}$$

with $k_n-k_n'\to 0$.

Now suppose $n^{2\beta_0}X_n^2\to 0$ for a β_0 in $[0, \mu_\gamma]$. We know this is true at least for $\beta_0=0$. Choose a β in (β_0, μ_γ) and set $\beta_n=n^{\beta+\beta_0}$. These β_n satisfy (2.10.14) and thus, also, (2.10.16). Thus $n^{\beta}X_n\to 0$ for every $\beta<\mu_\gamma$; but since μ_γ can be chosen as any number less than $1/2-\gamma-2\varepsilon_1$ (see (2.6.2)), the assertion of the theorem holds.

3. Asymptotic normality of the proposed procedure

To obtain the following asymptotic normality result for the procedure proposed in (2.6.3) we use a one-dimensional version of Theorem 2.2, Fabian [7].

3.1. *Asymptotic normality theorem*

If Assumption 2.8 *holds, then* $n^{1/2-\gamma}(X_n-\theta)$ *is asymptotically normal with*

(i) $\begin{cases} \textit{mean}=0, & \\ \textit{variance}=[6I(g)M^2C^2]^{-1}, & \textit{if } \gamma=1/4, \end{cases}$

(ii) $\begin{cases} \textit{mean}=0, & \textit{if } \gamma=1/6 \\ \textit{variance}=[(16/3)I(g)M^2C^2]^{-1}, & \textit{and } D^3f(\theta)=0, \\ \textit{mean}=-(Q/128)^{1/3}, & \textit{if } \gamma=1/6 \\ \textit{variance}=(Q/2^{11/2})^{2/3}, & \textit{and } D^3f(\theta)\neq 0, \end{cases}$

where $Q=3D^3f(\theta)M^{-3}I^{-1}(g)$.

PROOF. Assume without loss of generality that $\theta=0$. Suppose Assumption 2.8 holds. As in I, proof of Theorem 3.1, iii, use (2.8.1) and the Schwarz inequality to obtain $\limsup [f^{c_n}(X_n)]^{-1}\Psi_n[f^{c_n}(X_n)]\leqq I(g)$. This, along with (2.11.1) gives

$$[f^{c_n}(X_n)]^{-1}\Psi_n[f^{c_n}(X_n)]\to I(g)\ , \tag{3.1.1}$$

and from (2.10.6)

$$N_n=n^{-1}(\alpha_nX_n+\xi_nc_n^{-1+1/2r})\lambda_n\ , \tag{3.1.2}$$

with $\lambda_n=D_n[f^{c_n}(X_n)]^{-1}\Psi_n[f^{c_n}(X_n)]+(\log n)^{-1+\epsilon_1}\kappa_n$, where in the proof of Theorem 2.10, it was shown that $\kappa_n\to 1$. So by (3.1.1) and (2.7.1) we have

$$\lambda_n\to(2M)^{-1}\ . \tag{3.1.3}$$

Denoting conditional variance, given $\mathcal{F}_n$, by $\mathrm{Var}^{\mathcal{F}_n}$, we have

$$\mathrm{Var}^{\mathcal{F}_n}[h_n(Y_n)]=\int h_n^2(t+f^{c_n}(X_n))dG(t)-\Psi_n^2[f^{c_n}(X_n)]\to I(g)$$

by (2.8.1) and since $\Psi_n[f^{c_n}(X_n)]\to 0$. Therefore

$$D_n^2\,\mathrm{Var}^{\mathcal{F}_n}[h_n(Y_n)]\to(2M)^{-2}I^{-1}(g)\ . \tag{3.1.4}$$

Now consider $\mathrm{Var}^{\mathcal{F}_n}[(\log n)^{-1+\epsilon_1}\tilde{Y}_n]$. If y_n in (2.6.4) are $(\log n)^{1-2\epsilon_1}$, this variance is bounded by $(\log n)^{-2\epsilon_1}$. On the other hand, if $y_n=n^{\epsilon_1}$ then G has finite second moment, say σ^2, and $\mathrm{Var}^{\mathcal{F}_n}\tilde{Y}_n\leqq \mathrm{E}^{\mathcal{F}_n}\tilde{Y}_n^2\leqq \mathrm{E}^{\mathcal{F}_n}Y_n^2\leqq\sigma^2+[f^{c_n}(X_n)]^2$. So on the set $\{X_n\to 0\}$, $\limsup \mathrm{Var}^{\mathcal{F}_n}\tilde{Y}_n\leqq\sigma^2$. In either case then we have

$$\mathrm{Var}^{\mathcal{F}_n}[(\log n)^{-1+\epsilon_1}\tilde{Y}_n]\to 0\ . \tag{3.1.5}$$

The random variables $h_n(Y_n)$ and $\tilde{Y}_n$ are not independent, but by the Schwarz inequality it follows from (3.1.4) and (3.1.5) that

$$(nc_n)^2\,\mathrm{E}^{\mathcal{F}_n}(U_n-N_n)^2\to I^{-1}(g)(2M)^{-2}\ . \tag{3.1.6}$$

Now we set $Z_n=nc_n(U_n-N_n)$ and suppose r is in R^+. By (2.6.1), $Z_n=\mathcal{O}_u((\log n)^{\epsilon_0}n^{2\epsilon_1})$. So $\{Z_n^2>rn\}$ is eventually empty and

$$\mathrm{E}\,Z_n^2\chi_{\{Z_n^2\geqq rn\}}\to 0\ . \tag{3.1.7}$$

Writing X_{n+1} as $(X_n-N_n)-(U_n-N_n)$ we obtain $X_{n+1}=(1-n^{-1}\alpha_n\lambda_n)X_n-n^{-1}\cdot c_n^{-1}Z_n-n^{-1}c_n^{-1+1/(2r)}\xi_n\lambda_n$. Using this, (3.1.3), (3.1.6), (3.1.7), and the measurability properties of α_n, λ_n, and ξ_n we obtain the desired result by applying Theorem 2.2, Fabian [7] with U_n in Theorem 2.2 replaced by

X_n here, Γ_n by $\alpha_n\lambda_n$, V_n by Z_n, ϕ_n by $-C_n^{-1}$, and T_n by $-C_n^{-1+1/(2\gamma)}\xi_n\lambda_n$.

For the case $\gamma=1/4$, by (2.8.2), (2.10.5), and (3.1.3) we have, $\Phi=-C^{-1}$ and $T=0$. Similarly for the case when $\gamma=1/6$ and $D^3f(0)=0$. Finally, if $\gamma=1/6$ and $D^3f(0)\neq 0$, Φ is $-[(2/3)[D^3f(0)]^2I(g)]^{1/6}$ and T is $-(6M)^{-1}[(3/2)[D^3f(0)]I^{-1}(g)]^{1/3}$.

In all cases, $\Gamma=1$, $\alpha=1$, $\beta=\beta_+=1-2\gamma$, and by (3.1.6), $\Sigma=I^{-1}(g)\cdot(2M)^{-2}$.

4. The main result

In this section we state and prove the main result given in Theorem 4.1. The procedure given includes methods of estimation for all unknown parameters. Only 2.3 and 2.4 are assumed to hold.

4.1. *Theorem*

Suppose Assumptions 2.3 *and* 2.4 *hold with* $\mathscr{F}_n$ *as defined below. Let* $\{k_l\}$ *be an increasing sequence of positive integers such that* $l/k_l\to 0$. *Suppose* $\{U_l\}$ *and* $\{V_l\}$ *are sequences of random variables such that with*

$$\mathscr{F}_n=\mathscr{F}(\{X_1, Y_1,\cdots, Y_{n-1}\}\cup\{U_l;\, k_l<n\}\cup\{V_l;\, k_l<n\})\ ,$$

we have

$$\text{(4.1.1)}\qquad \begin{cases} \mathrm{E}^{\mathscr{F}_{k_l}}U_l=(2d_l)^{-2}(f^{d_l})^{d_l}(X_{k_l})\ , \\ \mathrm{E}^{\mathscr{F}_{k_l}}(U_l-\mathrm{E}^{\mathscr{F}_{k_l}}U_l)^2=\mathcal{O}_u(d_l^{-4})\ , \\ \mathrm{E}^{\mathscr{F}_{k_l}}V_l=(2d_l)^{-3}((f^{d_l})^{d_l})^{d_l}(X_{k_l})\ , \\ \mathrm{E}^{\mathscr{F}_{k_l}}(V_l-\mathrm{E}^{\mathscr{F}_{k_l}}V_l)^2=\mathcal{O}_u(d_l^{-6})\ , \end{cases}$$

with d_l *of the form*

$$\text{(4.1.2)}\qquad d_l=dl^{-\delta}\ ,\qquad d\ \text{in}\ R^+,\ 0<\delta<1/6\ .$$

Then the sequence $\{X_n\}$ *as defined in* 4.2 *below converges to* θ *and* $t_n^{1/2-\gamma}(X_n-\theta)$ *is asymptotically normal with mean and variance as given in Theorem* 3.1, (i) *and* (ii), *where* $2t_n=2n+7$ *card* $\{l;\, k_l<n\}$ *is the number of observations needed to construct* X_n.

4.2. *The procedure*

(i) *Estimation of* $D^2f(\theta)$:

Set $\bar{U}_n$ equal to the arithmetic mean of all U_l with $k_l<n$. Then set

$$\text{(4.2.1)}\qquad u_n=(0\vee\bar{U}_n)\ .$$

(ii) *Estimation of* $D^3 f(\theta)$:

Set v_n equal to the arithmetic mean of all V_l with $k_l < n$.

(iii) *Estimation of* $I(g)$:

Let $\beta_0 > 0$ and choose $\varepsilon_n \geqq (\log n)^{-\beta_0}$. Then this estimation is carried out precisely as it is in (I4.2.b), that is by a sequence $\{w_n\}$ with

$$w_n = \int (h_n^\circ)^2 dG_{n-1} \tag{4.2.2}$$

where G_n is the empirical distribution function of $Y_1, Y_2, \cdots, Y_n$, and h_n° is defined by

$$h_{n+1}^\circ(t) = -\frac{(G_n^{\delta_n})^{\Delta_n}(t)}{2\delta_n G_n^{\Delta_n}(t)} \chi_{(\varepsilon_n, +\infty)}(G_n^{\Delta_n}(t)) \tag{4.2.3}$$

for all t in $T_n = \{(2j-1)\Delta_n;\ j = 0, 1, -1, \cdots\}$ and let h_{n+1}° be constant on the intervals $((2j-2)\Delta_n, 2j\Delta_n]$, where Δ_n and δ_n are sequences of positive numbers such that $\Delta_n \to 0$, $\varepsilon_n \Delta_n^{-1} \to 0$, $\delta_n \varepsilon_n^{-1} \to 0$ and $n^{-r}\delta_n^{-1}\varepsilon_n^{-1} \to 0$ for an $r < 1/2$.

(iv) *The sequence* C_n:

Set

$$C_n = \begin{cases} [((3/2)v_n^{-2}w_n^{-1})^{1/6} \vee (\log n)^{-\varepsilon_0}] \wedge (\log n)^{\varepsilon_0}, & \text{if } \gamma = 1/6 \text{ and } D^3 f(g) \neq 0 \\ C, & \text{otherwise} \end{cases} \tag{4.2.4}$$

with ε_0 as in (2.6.1), C as in (2.8.2), v_n as in (4.2.ii) and w_n as in (4.2.2).

(v) *The sequence* D_n:

Set

$$D_n = (2u_n w_n)^{-1} \wedge n^{\varepsilon_1}. \tag{4.2.5}$$

(vi) *The functions* h_n:

Choose h_n° to satisfy (4.2.3) with $\varepsilon_n \geqq n^{-\beta_1}$ and $0 < \beta_1 < 1/2 - \gamma - 2\varepsilon_1$, with γ as in Assumption 2.3 and ε_1 as in (2.6.1). Then set

$$h_n(t) = \begin{cases} (1/2(h_n^\circ(t) - h_n^\circ(-t)) \vee 0) \wedge (n^{\varepsilon_1}\chi_{(-n,n)}(t)), & \text{for } t \geqq 0, \\ -h_n(-t), & \text{for } t < 0. \end{cases} \tag{4.2.6}$$

(vii) *The sequence* X_n:

The recursion relation for X_n is given in (2.6.3).

PROOF OF THEOREM 4.1. We shall prove the theorem by verifying Assumptions 2.3, 2.4, 2.6, 2.7, and 2.8.

First, Assumptions 2.3 and 2.4 are assumed to hold in the theorem. The measurability conditions on C_n, D_n, and h_n and condition (2.6.1) are obvious from their definitions. Relation (2.6.3) holds by assumption. Thus Assumption 2.6 holds, and by Theorem (2.10), $(\log n)^\beta(X_n-\theta)\to 0$, for every $\beta>0$.

To show that u_n converges to $D^2f(\theta)$, it suffices to show that $\bar{U}_n$ does. Let $W_l=U_l-\mathrm{E}^{\mathcal{F}_{k_l}}U_l$. Then W_l is an orthogonal sequence, $\sum_{l=1}^{\infty}(\log l)^2 \cdot l^{-2}\,\mathrm{E}\,W_l^2 \leqq C_1 d^{-4}\sum_{l=1}^{\infty}(\log l)^2 l^{-2+4\delta}<+\infty$ for a C_1 in R^+. So by Theorem 33.1.B.ii, Loève [12], we have $l^{-1}\sum_{j=1}^{l}W_j\to 0$. Also $\mathrm{E}^{\mathcal{F}_{k_l}}U_l=D^2f(X_{k_l}+\nu_l)$, where $|\nu_l|\leqq 2d_l$. So eventually, depending on ω, we obtain using Assumption 2.3 that $\mathrm{E}^{\mathcal{F}_{k_l}}U_l\to D^2f(\theta)$ and $\bar{U}_n\to D^2f(\theta)$. The convergence of w_n to $I(g)$ follows from Theorem 2.2, Fabian [10]. Verification of the assumptions of this theorem are given in (I4.3.ii). Therefore Assumption 2.7 holds, and by our Theorem 2.11, $n^\beta(X_n-\theta)\to 0$, for every $0<\beta<1/2-\gamma-2\varepsilon_1$.

Finally, (2.8.1) follows from Extension 2.3, Fabian [10]. Details and verification of the assumptions of this extension are given in (I4.3.iii). The convergence of v_n to $D^3f(\theta)$ follows by an argument similar to that used to show $\bar{U}_n\to D^2f(\theta)$. Therefore Assumption 2.8 holds, and by our Theorem 3.1, X_n has the properties asserted in Theorem 4.1 since $t_n/n\to 1$ because $l/k_l\to 0$.

THE OHIO STATE UNIVERSITY*

REFERENCES

[1] Abdelhamid, S. N. (1973). Transformation of observations in stochastic approximation, *Ann. Statist.*, **1**, 1158-1174.

[2] Anbar, D. (1973). On optimal estimation methods using stochastic approximation procedures, *Ann. Statist.*, **1**, 1175-1184.

[3] Blum, J. R. (1954). Approximation methods which converge with probability one, *Ann. Math. Statist.*, **25**, 737-744.

[4] Dubins, L. E. and Freedman, D. A. (1965). A sharper form of the Borel-Cantelli lemma and the strong law, *Ann. Math. Statist.*, **36**, 800-807.

[5] Fabian, V. (1960). Stochastic approximation methods, *Czechoslovak Math. J.*, **10**(85), 123-159.

[6] Fabian, V. (1964). A new one-dimensional stochastic approximation method for finding a local minimum of a function, *Trans. Third Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*, Czechoslovak Academy of Sciences, Prague, 85-105.

[7] Fabian, V. (1968). On asymptotic normality in stochastic approximation, *Ann. Math. Statist.*, **39**, 1327-1332.

[8] Fabian, V. (1971). Stochastic approximation, *Optimizing Methods in Statistics* (ed. S. S. Rustagi), Academic Press, New York, 439-470.

[9] Fabian, V. (1973a) (Referred to in the text as I). Asymptotically efficient stochastic approximation; the RM case, *Ann. Statist.*, **1**, 486-495.

* Now at University of Denver.

[10] Fabian, V. (1973b). Estimation of the derivative of the logarithm of a density, *Ann. Statist.*, **1**, 557-561.

[11] Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.*, **23**, 462-466.

[12] Loève, Michel (1963). *Probability Theory*, Third Edition, D. van Nostrand Company, Inc.

[13] Robbins, H. and Monro, S. (1951). A stochastic approximation method, *Ann. Math. Statist.*, **22**, 400-407.

[14] Venter, J. H. (1967). An extension of the Robbins-Monro procedure, *Ann. Math. Statist.*, **38**, 181-190.