Václav Fabian

Stochastic approximation methods

Czechoslovak Mathematical Journal, Vol. 10 (1960), No. 1, 123–159

Persistent URL: http://dml.cz/dmlcz/100396

Чехословацкий математический журнал, т. 10 (85) 1960, Прага

STOCHASTIC APPROXIMATION METHODS

VÁCLAV FABIAN, Praha

(Received March 25, 1959)

Some modifications of known approximation procedures are considered and general theorems are proved, which make possible the study of their convergence.

1. Introduction. Stochastic approximation methods deal with the problem of approximating a point of the q-dimensional Euclidian space E_q at which a function f acquires its minimum (or maximum or at which the value of f is equal to a predetermined number). Such problems are of great importance, especially in connection with practical problems of finding optimum conditions for concrete chemical and physical processes, where we usually can for every point x in E_q (which represents some fixed conditions of the process considered) determine or at least estimate the number $f(x)$ (which describes the „quality" of the process with conditions characterised by x). Often the only further knowledge that we have about f, is that f has some very general properties (has a bounded second derivative etc.). The approximation process starts with a point (or random vector) X_1 in E_q and constructs successively a sequence of random vectors $X_1, X_2, \ldots$. The first n members $X_1, X_2, \ldots, X_n$ being observed, the sequence proceeds in such a direction Y_n that it can be expected that f decreases on the segment $[X_n, X_n + aY_n]$ at least for small a. Then the length α_n of the n-th step is chosen and X_{n+1} is defined to be $X_n + \alpha_n Y_n$.

Methods for attaining optimum conditions were proposed even before the origin of stochastic approximation methods, but only intermediate steps were studied in the search for a minimum (factorial and other designs, the method of G. E. P. BOX and K. B. WILSON ([4], 1951)).

Although till now the majority of papers are concerned with the one-dimensional case, it seems that the multidimensional case is of an incomparably greater importance from the practical point of view. Indeed, in the one-dimensional case (and much less in the two-dimensional) a graphic description of data makes possible subjective considerations, which may in some cases be more efficient than a general objective scheme. However, in the three- and more-dimensional case, the possibility of a graphic representation breaks down and the systematic approach is also inconvenient, since the number of points of a reasonably dense net in the domain of f increases geometrically with the dimension. If we try to represent the function considered by a polynomial, then, if we have to find an extremal point, the degree of the polynomial must be greater than one and usually the representation does not lead to a practical reduction of the problem. On the other hand, if stochastic approximations are used, then, under certain conditions and in a sense to be specified later in section 9, an increase in the number of dimensions from $q - 1$ to q increases the number of observations by a factor of $\frac{q+1}{q}$ (see (9.2)). Hence, and from practical experience, it seems that the use of multidimensional stochastic approximation can lead to a substantial increase in the efficiency of experimental work e. g. in chemistry, engineering, zoology, medicine and so on. Moreover it seems that some results in multidimensional stochastic approximation are new also in the particular case when the values of the function considered can be determined precisely without any random error in which case it deals with a problem in numerical analysis rather than in probability. In this sense, the stochastic approximation methods are related to more special methods (so called methods of the steepest descent, see e. g. [12]) and seem to be more fit than they for use in constructing an automatic optimizer (see [10]).

To fix the ideas let

(1.1) $$X_{n+1} = X_n + \alpha_n Y_n ,$$

where X_n, Y_n are q-dimensional random vectors, α_n are random variables, let us write $\mathscr{X}_n = [X_1, \ldots, X_n]$ and let us denote by $\mathbf{M}_n(\mathscr{X}_n)$ the conditional expectation $\mathbf{E}_{\mathscr{X}_n} Y_n$ of Y_n given $\mathscr{X}_n$.

The pioneering paper of H. Robbins and S. Monro ([15], 1951) deals with the (one-dimensional, i. e. $q = 1$) problem of finding a root of an equation $R(x) = 0$. Under somewhat stronger conditions than are those of the following theorem, Robbins and Monro proved the convergence of X_n to the solution Θ in probability; under the conditions of the following theorem, the convergence with probability one was proved by J. R. Blum [1] in 1954.

(1.2) **Theorem.** (Robbins-Monro method.) *Suppose that R is a function defined on $E_1 = (-\infty, +\infty)$, that*

(1.2.1) $$\sup_{-k<x-\Theta<-\frac{1}{k}} R(x) < 0 , \quad \inf_{\frac{1}{k}<x-\Theta<k} R(x) > 0$$

for a (unknown) number Θ and every natural number k and that there exist constants A, B such that

(1.2.2) $$|R(x - \Theta)| < A\,|x - \Theta| + B$$

for all x.

Now if a_n is a sequence of positive numbers such that

$$\sum_{n=1}^{\infty} a_n = +\infty, \quad \sum_{n=1}^{\infty} a_n^2 < +\infty, \tag{1.2.3}$$

if in (1.1) $\alpha_n = a_n$, if

$$\mathbf{M}_n(\mathscr{X}_n) = -R(X_n) \tag{1.2.4}$$

and if

$$\mathbf{E}_{\mathscr{X}_n}(Y_n - M_n(\mathscr{X}_n))^2 \leqq \sigma^2 \tag{1.2.5}$$

for a suitable σ and for every natural number n, then the sequence X_n converges to the point Θ with probability one.

We observe that (1.2.4) states that Y_n is an unbiased estimate of $-R(X_n)$ and that the sequence of (conditional) variances of Y_n is bounded. The intuitive reason for defining X_{n+1} to be $X_n + a_n Y_n$ is that $E_{\mathscr{X}_n} a_n Y_n = -a_n R(X_n)$ is by (1.2.1) positive and negative for $X_n < \Theta$ and for $X_n > \Theta$ respectively.

In 1952 J. Kiefer and J. Wolfowitz [14] solved in a analogous way the problem of finding a maximum of a function R defined on E_1 and proved that under suitable conditions their scheme converges in probability. Again J. R. Blum [1] has weakened the conditions and proved the convergence with probability one; this result is recapitulated in the following

(1.3) **Theorem.** (Kiefer-Wolfowitz method.) *Suppose that R is a function defined on E_1, that*

$$\inf_{-k < x - \Theta < -\frac{1}{k}} \underline{D}R(x) > 0, \quad \sup_{\frac{1}{k} < x - \Theta < k} \overline{D}R(x) < 0 \tag{1.3.1}$$

for a (unknown) number Θ and every natural number k, where $\underline{D}f(x)$ and $\overline{D}f(x)$ denote the lower and upper derivative respectively of the function f at the point x. Suppose that there exist constants A, B such that

$$|R(x+1) - R(x)| < A\,|x - \Theta| + B \tag{1.3.2}$$

for all x.

Let a_n, c_n be two sequences of positive numbers,

$$c_n \to 0, \sum a_n = +\infty, \quad \sum \frac{a_n^2}{c_n^2} < +\infty, \tag{1.3.3}$$

let $\alpha_n = a_n$,

$$\mathbf{M}_n(\mathscr{X}_n) = \frac{1}{2c_n}\,[R(X_n + c_n) - R(X_n - c_n)], \tag{1.3.4}$$

and

$$\text{(1.3.5)} \qquad \mathbf{E}_{\mathscr{X}_n}(Y_n - \mathbf{M}_n(\mathscr{X}_n)) \leqq \frac{\sigma^2}{2c_n^2} \;{}^{1)}$$

for a suitable σ and for every natural number n.

Then $\mathscr{X}_n$ converges to Θ whit probability one.

In 1954 Blum [2] generalized the one-dimensional results of Robbins, Monro and Kiefer, Wolfowitz to their multidimensional analogues. However it seems to us that the conditions of the multidimensional Blum's analogue to the Robbins-Monro method are too strong and that the second Blum's method desribed in the following theorem is of a considerably greater importance. Before stating the theorem we introduce the following notations.

For a vector x we denote by $x^{(i)}$ the i-th component of x, for a matrix M we denote by $M^{(ij)}$ the element of the i-th row and j-th column; further we denote $\|x\| = \sqrt{\sum_{i=1}^{q} [x^{(i)}]^2}$, $\|M\| = \sup_{\|x\|=1} \|Mx\|$. By Δ_i we denote the vector satisfying $\Delta_i^{(j)} = 0$ for $j \neq i$ and $\Delta_i^{(i)} = 1$. If f is a function on E_q, then by the symbols $Df(x)$ and $D_2f(x)$ we mean the vector and matrix such that $D^{(i)}f(x) = [Df(x)]^{(i)} = \frac{\partial}{\partial x^{(i)}} f(x)$ and $D_2^{(ij)}f(x) = [D_2f(x)]^{(ij)} = \frac{\partial^2}{\partial x^{(i)}\, \partial x^{(j)}} f(x)$.

(1.4) **Theorem.** (Blum's method.[2]) *Suppose that R is a function defined on $\mathbf{X} = E_q$, that $DR(x)$ and $D_2R(x)$ exist for all $x \in \mathbf{X}$, that*

$$\text{(1.4.1)} \qquad R(\Theta) = 0, \inf\{R(x); \|x - \Theta\| > \varepsilon\} > 0, \inf\{\|DR(x)\|; \|x - \Theta\| > \varepsilon\} > 0^{3)}$$

for a $\Theta \in \mathbf{X}$ and every $\varepsilon > 0$ and that

$$\text{(1.4.2)} \qquad \|D_2R(y)\| \leqq 2K$$

for a suitable constant K.

If a_n, c_n are positive numbers such that

$$\text{(1.4.3)} \qquad c_n \to 0\,, \sum a_n = +\infty, \sum a_n c_n < +\infty\,, \sum \frac{a_n^2}{c_n^2} < +\infty\,,$$

---

[1]) Usually Y_n is supposed to be $\frac{Y_n^+ - Y_n^-}{2c_n}$, where Y_n^+ and Y_n^- are estimates of $R(X_n + c_n)$ and $R(X_n - c_n)$ respectively,

$$E_{\mathscr{X}_n}(Y_n^+ - R(X_n + c_n))^2 \leqq \sigma^2\,, \quad E_{\mathscr{X}_n}(Y_n^- - R(X_n - c_n))^2 \leqq \sigma^2\,.$$

[2]) We change inessentially the original theorem by considering the function $R = -M$, where M is the function considered by Blum [2].

[3]) Hence R has its unique minimum at Θ.

if in (1.1) $\alpha_n = a_n$,

$$\mathbf{M}_n^{(i)}(\mathscr{X}_n) = -\frac{1}{c_n}\left(R(X_n + c_n\Delta^{(i)}) - R(X_n)\right), \tag{1.4.4}$$

$$\mathbf{E}_{\mathscr{X}_n}(Y_n - \mathbf{M}_n(\mathscr{X}_n))^2 \leqq \frac{\sigma^2}{c_n^2}, \tag{1.4.5}$$

then X_n *converges to* Θ *with probabitity one.*

The reader should note the analogy between (1.2.1), (1.3.1) and (1.4.1). On the other hand (1.4.2) is much stronger than (1.2.2) and (1.3.2). However in the one-dimensional case there are only two possible directions for the move from X_n to X_{n+1}; in the multidimensional case there are uncountably many directions and only q directions are examined by Y_n; this is the reason for stronger conditions on $DR(x)$.

From further studies on the convergence we mention the paper by A. Dvoretzky ([9], 1956), in which the problem of stochastic approximation was attacked with considerable generality, making it possible especially to obtain in a unified way all the previous results concerning convergence properties — both in mean square and with probability one — in the one-dimensional case.

In 1958 H. Kesten [13] proposed a modification of the Robbins-Monro procedure substituting the definition $\alpha_n = a_n$ by the definition $\alpha_1 = a_1$, $\alpha_2 = a_2$, $\alpha_n = a_{m+2}$, where m denotes the number of changes of sign in the sequence $Y_1, Y_2, \ldots, Y_{n+1}$, i. e. $m = \frac{1}{2}\sum_{i=1}^{n-2} |\text{sign}\, Y_{i+1} - \text{sign}\, Y_i|$. The intuitive reason for the modification is that small m indicates that $|X_n - \Theta|$ is large and that it is unreasonable to diminish α_n. Under the additional assumption that a_n is a nonincreasing sequence and under some additional weak assumptions on Y_n, Kesten proved the convergence with probability one to Θ of the modified Robbins-Monro procedure. He studied also the Kiefer-Wolfowitz procedure but was unable to prove that its analogous modification preserves its convergence to Θ. He proved this convergence only after some further changes in assumptions especially after replacing the condition $c_n \to 0$ by $c_n = \text{const.}$; however, in this case X_n does not in general converge to the point (if it exists) at which R acquires its maximum.

In 1958 Václav Dupač [8] devised an essentially new method for solving simultaneous equations $R_i(x) = 0$ $(i = 1, \ldots, q)$ under the assumption that R_i are linear functions.

In addition to the construction of new approximation methods and the proof of their convergence, the speed of this convergence, at least asymptotically, was studied in a number of papers among which the first was that by K. L. Chung [11]. The method of Chung, who deals with the process of Robbins-Monro only, was applied to the study of the Kiefer-Wolfowitz pro-

procedure independently by C. DERMAN ([5], 1956) and Václav Dupač [7], 1957) with partially overlapping results. Václav Dupač studied also—using Chung's method — the asymptotic speed of convergence of his above-mentioned multidimensional stochastic procedure ([8], 1958). A very general and fruitful study in this direction concerning both one- and multidimensional cases was published in 1958 by JEROME SACKS [16], who used succesfully another method of proof than Chung. All results of this kind are of great importance for their consequences for the choice of eligible constants in the schemes studied. The best choice (unique minimax in the non-asymptotic sense) of eligible constants in a special case of Robbins-Monro procedure was found by Dvoretzky in the already cited paper [9].

In the present paper we propose two modifications of the known procedures and study their convergence with probability one. In order not to interfer with the convergence property, the modifications which lead to a weakening of the conditions concerning the function R require stronger conditions for the estimates of the values of R. However these strengthened conditions are still rather general since they are satisfied if, roughly speeking, all errors of the estimates of the values of R used in the approximation process are continuous and equally distributed (see Theorems (4.3), (4.4), (8.4), (8.5), sections (6.2) and (7.2)).

The recurrence relation (1.1) for the Robbins-Monro procedure can be rewritten in the form

$$X_{n+1} = X_n + a_n |Y_n| \operatorname{sign} Y_n ,$$

where Y_n is an estimate of $- R(X_n)$. Hence we see that the direction of the n-th move of the approximation process is choosen to be sign Y_n and the length of the move is chosen to be $a_n |Y_n|$. This choice will be reasonable if large values of $|Y_n|$ can be expected for large $|X_n - \Theta|$, but this is not guaranteed by the assumptions of the Robbins-Monro method. Thus if e. g. $R(X) = Xe^{-X^2}$ then assumptions (1.2.1) and (1.2.2) are satisfied for $\Theta = 0$, the Robbins-Monro procedure still converges to 0, but it behaves unsatisfactory from the practical point of view. Indeed it makes small corrections for $|X_n - \Theta|$ large and large corrections if $|X_n - \Theta|$ is small. If we determine the length of the n-th move to be a_n instead of $a_n |Y_n|$, we get a procedure much less charged by this inconveniency (and free of it if there is no error in observations). Moreover the above-mentioned weakening of conditions imposed on R consists in omitting (1.2.2). That (1.2.2) cannot be omitted without the modification of the procedure, follows from the following example (see Dvoretzky [8]): Let $R(x) = |x| x$, $Y_n = R(X_n)$ (i. e. there is no error in observation), $a_n = \frac{1}{n}$, $X_0 = 3$; then $X_2 = 3 - 3^2 = - 6$, $X_3 = - 6 + \frac{6^2}{2} = 12, \ldots$ and it is easily verified that $|X_n| \to + \infty$ if the original approximation scheme is used.

On the other hand for the above described modification we have $X_0 = 3$, $X_1 = 3 - 1 = 2$, $X_2 = 2 - \frac{1}{2} = \frac{3}{2}$, $X_3 = \frac{3}{2} - \frac{1}{3} = \frac{7}{6}, \ldots$ and $X_n \to 0$.

The situation in the case of the Kiefer-Wolfowitz method is analogous. Here the length of the n-th step is $\frac{a_n}{2c_n} |Y_n|$ which again seems not to be reasonable unless a further assumption (here that of concavity) concerning R is satisfied. In the general case we propose to modify the procedure by taking $\frac{a_n}{2c_n}$ for $\frac{a_n}{2c_n} |Y_n|$, so that (1.1) changes to

$$X_{n+1} = X_n + \frac{a_n}{2c_n} \operatorname{sign} Y_n \, .$$

In the multidimensional case we study the modification consisting in replacing $Y_n^{(i)}$ by sign $Y_n^{(i)}$ $(i = 1, \ldots, q)$. As the proposed modification of the determination of the length of the n-th step of the process makes possible the omission of the condition (1.2.2) in the case of the Robbins-Monro method, it enables us to omit the condition (1.3.2) in the case of the Kiefer-Wolfowitz method and to weaken the condition (1.4.2) in the multidimensional case of Blum (only however, if conditions on Y_n are strengthened).

The second modification is motivated by the fact that in the search for a minimum of a function by the method of Blum we need at least $q + 1$ observations for determining the direction at each step. Since we never know the optimum length of the move, it seems to be unreasonable, especially if q is large, to examine only one length. We propose to determine the length α_n in the following way: If X_n and Y_n are observed, take observations V_j (independent of X_n, Y_n) of $R(X_n + jaY_n)$ for $j = 1, 2, \ldots$ until $V_1 > V_2 > \ldots > V_{j-1}$ and put $\alpha_n = ja$ if $V_1 > V_2 > \ldots > V_{j-1} > V_j \leqq V_{j+1}$.

Thirdly we study the behaviour of the sequence X_n if the assumption (1.4.1) is not required. It can be shown in this case that $f(X_n)$ is a convergent sequence which behaves as if the sequence X_n converges to a zero-point of the derivative of f (see Note (5.3)). It is paradoxal that we have not succeeded in proving that this must be a local minimum, but it seems that this is a weakness of our methods of proofs rather than a deficiency of the approximation methods.

Concerning the ordering of the paper, section 2 introduces some notations and assumptions, sections 3 and 4 deal with the modification of α_n mentioned above. The reader interested only in the case, in which α_n are numbers, can omit these sections except Theorem (4.1). Section 5 contains basic convergence theorems; in Note (5.3) the interpretation of results is discussed. Section 6, 7 and 8 contains proofs of Theorems (1.1), (1.2) and (1.3) — and their generalisations — respectively. Some concluding remarks are made in section 9.

2. Some notations and basic assumptions. Let q be an integer and $\mathbf{X} = E_q$ the q-dimensional Euclidean space. If x, y are in $\mathbf{X}$, we denote by $\langle x, y \rangle$ the inner product $\sum_{i=1}^{q} x^{(i)} y^{(i)}$ of x and y. The norm $\|x\| = \sqrt{\langle x, x \rangle}$ of a vector x and the norm of a matrix were defined in the preceding section.

Let $(\Omega, \mathscr{F}, P)$ be a probability space. By random variables we mean measurable transformations from Ω to E_1, by random vectors we mean measurable transformations from Ω to $\mathbf{X}$. If X is a random vector, then we denote by $X^{(i)}$ the random variable defined by the relation $X^{(i)}(\omega) = [X(\omega)]^{(i)}$, by $\mathbf{E}X$ (expectation of X) the vector defined by the relation $[\mathbf{E}X]^{(i)} = \int X^{(i)} \, \mathrm{d}P$, if these integrals have a meaning for every $i = 1, 2, \ldots, q$. By $\mathbf{D}X$ we denote the $q \,.\, q$ (covariance) matrix the element $\mathbf{D}^{(ij)}X$ of which equals $\mathbf{E}X^{(i)}X^{(j)}$. Concerning equalities, inequalities and convergence of random vectors or variables, they are always meant with probability one.

In the sequel we shall deal with a function f satisfying

(2.1) **Assumption.** f is a non-negative real valued function defined on $\mathbf{X}$, $D_2 f(x)$ exists for every $x \in \mathbf{X}$ and $\|D_2 f(x)\| \leqq 2K$ for a number K and every $x \in \mathbf{X}$.

For simplicity we shall write $D(x) = Df(x)$; if Assumption (2.1) is satisfied, then by Taylor's Theorem we get

$$f(x + y) \leqq f(x) + \langle y, D(x) \rangle + K \, \|y\|^2 \tag{2.1.1}$$

for every x, y in $\mathbf{X}$.

3. The choice of the random variables α_n. Given X_n and Y_n the random variable α_n determines the length of the move from X_n in the direction determined by Y_n. Let a be a positive number; we shall suppose that α_n can acquire the values $a, 2a, \ldots$ only. This assumption is not essential, but removing it leads to complications of proofs or to results insufficiently general.

Let f be a function satisfying assumption (2.1). For $\omega \in \Omega$ we define two functions φ_ω, ψ_ω by the relations $\varphi_\omega(t) = f(X_n(\omega) + tY_n(\omega))$, $\psi_\omega(t) = \varphi_\omega(0) + t\varphi'_\omega(0) + t^2 K \, \| Y_n(\omega)\|^2$ for every $t \in E_1$. By Assumption (2.1) we have $\varphi'_\omega(0) = \langle Y_n(\omega), D(X_n(\omega))\rangle$, $\varphi_\omega(t) \leqq \psi_\omega(t)$. By $\tau^+(\omega)$ and $\tau^-(\omega)$ we denote the product ja where j is the largest principal such that the sequence $\varphi_\omega(a), \varphi_\omega(2a), \ldots, \varphi_\omega(ja)$ is increasing and decreasing respectively. From the two numbers $\tau^+(\omega)$ and $\tau^-(\omega)$ at least one is a; if the whole sequence $\{\varphi_\omega(ia)\}_{i=1}^{\infty}$ is increasing (decreasing), we put $\tau^+(\omega) = +\infty$ $(\tau^-(\omega) = +\infty)$.

Now let $P(\omega)$ be the system of such intervals $\langle (j-1)\, a, ja \rangle$ $\left(j = 1, 2, \ldots, \frac{\alpha_n(\omega)}{a}\right)$ for which $\varphi_\omega(ja) - \varphi_\omega((j-1)\, a) > 0$ and denote by $\alpha_n^+(\omega)$ the Lebesgue measure of the union $\bigcup P(\omega)$ of these intervals.

So we have defined three functions τ^+, τ^-, α^+ on Ω; clearly α_n^+ is a random variable. Since our aim is to minimize $\varphi_\omega(\alpha_n(\omega))$, we try to determine α_n so that α_n^+ would be small and that α_n would be in some sense not greater in the case $\tau^+(\omega) > a$ than in the case $\tau^-(\omega) > a$. In the next theorem we shall state conditions, under which α_n is at least as good as a random variable β independent of X_n and Y_n.

(3.1) **Theorem.** *Suppose there exist two numbers c_n, a_n and a non-negative random variable β, assuming values a, $2a$, ... only and such that*

(3.1.1) $$\mathbf{E}_{\mathscr{X}_n, Y_n} [\alpha_n^+]^2 \leqq c_n ,$$

(3.1.2) $$\mathbf{E}_{\mathscr{X}_n, Y_n} \beta = a_n , \quad \mathbf{E}_{\mathscr{X}_n, Y_n} \beta^2 \leqq c_n$$

and that for every ω in some subset Ω_0 of Ω

(3.1.3) $$\beta(\omega) < \tau^+(\omega) \Rightarrow \alpha_n(\omega) \leqq \beta(\omega)$$

and

(3.1.4) $$\alpha_n(\omega) < \tau^-(\omega) \Rightarrow \beta(\omega) \leqq \alpha_n(\omega)$$

and for every $\omega \in \Omega - \Omega_0$

(3.1.5) $$|\alpha_n(\omega) - c(\omega)| < \alpha_n^+(\omega) + a ,$$

where $c(\omega) \in E_1$ and

(3.1.6) $$\varphi_\omega(c(\omega)) \leqq \inf \{\varphi_\omega(t); t \geqq 0\} , \quad \varphi'_\omega(c(\omega)) = 0 .$$

Finally suppose that f satisfies Assumption (2.1). *Then*

(3.1.7) $$\mathbf{E}_{\mathscr{X}_n} f(X_{n+1}) \leqq f(X_n) + a_n < \mathbf{M}_n(X_n), D(X_n) > + + 11 c_n K \mathbf{E}_{\mathscr{X}_n} \|Y_n\|^2 .$$

Remark. Since in the theorem the index n is fixed, we can omit it in the symbols $\mathbf{M}_n$, $\mathscr{X}_n$, X_n, Y_n, α_n, α_n^+; for X_{n+1} we shall write $X + \alpha Y$ and $K(\omega)$ for $K \|Y(\omega)\|^2$. Before proving the theorem let us prove some lemmas.

(3.2) **Lemma.** *Suppose that f satisfies Assumption* (2.1.). *Then φ_ω has a continuous derivative*

(3.2.1) $$\varphi'_\omega(t) = \langle Y(\omega), \; D(X(\omega) + tY(\omega))\rangle$$

and a bounded (*by* $2K(\omega)$) *second derivative.*

For every t_1, t_2 we have

(3.2.2) $$\varphi_\omega(t_2) \leqq \varphi_\omega(t_1) + (t_2 - t_1)\, \varphi'_\omega(t_1) + (t_2 - t_1)^2\, K(\omega) ;$$

especially for every t

(3.2.3) $$\varphi_\omega(t) \leqq \psi_\omega(t) .$$

Proof. The conclusions follow from the assumption in a straightforward way.

(3.3) **Lemma.** *Suppose that f satisfies Assumption* (2.1). *Then there exists a function t_0 on Ω such that for every $\omega \in \Omega$ satisfying the condition* $\operatorname{Max}\{\tau^+(\omega), \tau^-(\omega)\} < +\infty$ *we have*

(3.3.1) $$\varphi'_\omega(t_0(\omega)) = 0 \,,$$

(3.3.2) $$\tau^+(\omega) = a \Rightarrow |t_0(\omega) - \tau^-(\omega)| < a$$

and

(3.3.3) $$\tau^-(\omega) = a \Rightarrow |t_0(\omega) - \tau^+(\omega)| < a \,.$$

Proof. Let $\omega \in \Omega$. If $\tau^+(\omega) = \tau^-(\omega) = a$, then $\varphi_\omega(a) = \varphi_\omega(2a)$ and thus there exists a $t_0(\omega) \in (\tau^-(\omega), \tau^-(\omega) + a) = (\tau^+(\omega), \tau^+(\omega) + a)$ so that (3.3.1) to (3.3.3) hold. If $\tau^+(\omega) \neq \tau^-(\omega)$ and $\tau^+(\omega) = a$ resp. $\tau^-(\omega) = a$, $\operatorname{Max}\{\tau^+(\omega), \tau^-(\omega)\} < < +\infty$, then

$$\varphi_\omega(\tau^-(\omega) - a) > \varphi_\omega(\tau^-(\omega)) \leqq \varphi_\omega(\tau^-(\omega) + a)$$

resp.

$$\varphi_\omega(\tau^+(\omega) - a) < \varphi_\omega(\tau^+(\omega)) \geqq \varphi_\omega(\tau^+(\omega) + a$$

so that again there exists a $t_0(\omega)$ satisfying (3.3.1) to (3.3.3).

(3.4) **Lemma.** *Suppose f satisfies Assumption* (2.1).

Then

(3.4.1) $$\tau^+(\omega) = a, \quad \tau^-(\omega) < +\infty \Rightarrow \varphi_\omega(\tau^-(\omega)) \leqq \operatorname*{Min}_{t \geqq a} \psi_\omega(t) + a^2 K(\omega) \,.$$

Proof. We discriminate two cases: (i) $\varphi'_\omega(0) \geqq 0$ and (ii) $\varphi'_\omega(0) < 0$.

(i) In this case ψ_ω increases in the interval $(0, +\infty)$, $\operatorname*{Min}_{t \geqq a} \psi_\omega(t) = \psi_\omega(a)$. From the definition of $\tau^-(\omega)$ we have $\varphi_\omega(\tau^-(\omega)) \leqq \varphi_\omega(a)$, according to (3.2.3) $\varphi_\omega(a) \leqq \psi_\omega(a)$: combining the three relations gives an inequality implying (3.4.1).

(ii) Denote $t_2 = \sup\{t'; \varphi'_\omega(t) < 0$ for every $t \in (0, t')\}$. From the assumption $\tau^-(\omega) < +\infty$ it follows that $t_2 < +\infty$. From the continuity of φ'_ω it follows that $t_2 > 0$ and $\varphi'_\omega(t_2) = 0$. From the definition of $\tau^-(\omega)$ it follows that $t_2 \in (0, \tau^-(\omega) + a)$ and we shall prove that

(3.4.2) $$\varphi_\omega(\tau^-(\omega)) \leqq \varphi_\omega(t_2) + a^2 K(\omega) \,.$$

Since the sequence $\varphi_\omega(a)$, $\varphi_\omega(2a), \ldots, \varphi_\omega(\tau^-(\omega))$ is decreasing, there exists a natural number j such that $\varphi_\omega(ja) \geqq \varphi_\omega(\tau^-(\omega))$ and $|j - t_2| < a$. Thus we get according to (3.2.2) $\varphi_\omega(\tau^-(\omega)) \leqq \varphi_\omega(ja) \leqq \varphi_\omega(t_2) + a^2 K(\omega)$ and (3.4.2) holds.

Now φ_ω has a second order derivative φ''_ω and $|\varphi''_\omega(t)| < 2K(\omega)$. Thus $|\varphi'_\omega(t) - - \varphi'_\omega(0)| < 2tK(\omega)$, which implies $t_2 > t_1 = \dfrac{-\varphi'_\omega(0)}{2K(\omega)}$ and hence $\varphi_\omega(t_2) \leqq \leqq \varphi_\omega(t_1)$. On the other hand it is easy to see that $\psi_\omega(t_1) = \operatorname*{Min}_{t \geqq a} \psi_\omega(t)$. Combining our results, we get $\varphi_\omega(\tau^-(\omega)) \leqq \varphi_\omega(t_2) + a^2 K(\omega) \leqq \varphi_\omega(t_1) + a^2 K(\omega) \leqq \psi_\omega(t_1) +$

$+ a^2K(\omega) \leqq \underset{t \geqq a}{\text{Min}}\, \varphi_\omega(t) + a^2K(\omega)$. Thus (3.4.1) holds in the case (ii) too, and the lemma is proved.

(3.5) **Lemma.** *Suppose all assumptions of Theorem* (3.1) *hold and put* $\tau = \text{Max}\,(\tau^-, \tau^+)$.

Then

$$\varphi_\omega(\alpha(\omega)) - \varphi_\omega(\tau(\omega)) \leqq K(\omega)\,\{8\,[\alpha^+(\omega)]^2\} \tag{3.5.1}$$

as soon as $\alpha(\omega) \geqq \tau(\omega)$ *and*

$$|\varphi_\omega(\beta(\omega)) - \varphi_\omega(\tau(\omega))| \leqq 2\beta^2(\omega)\,K(\omega) \tag{3.5.2}$$

as soon as $\tau(\omega) < +\infty$.

Proof. From the assumption $\alpha(\omega) \geqq \tau(\omega)$ it follows that $\tau^+(\omega)$ and $\tau^-(\omega)$ are finite. Remember that $\alpha^+(\omega)$ is the Lebesgue measure of the union $\bigcup P(\omega)$ of the system

$$P(\omega) = \{\langle ja, (j-1)\,a\rangle;\ \varphi_\omega(ja) - \varphi_\omega((j-1)\,a) > 0,\ j = 1, 2, \ldots,\ a \leqq ja \leqq \alpha\}\,.$$

Now $\bigcup P(\omega)$ can be written as a union of another system $B(\omega)$ of disjoint intervals $\langle c_i, d_i\rangle$ $(i = 1, 2, \ldots, k)$, where c_i, d_i are integral multiples of a,

$$\varphi_\omega(d_i) \geqq \varphi_\omega(c_{i+1})\,, \qquad i = 1, 2, \ldots, k-1 \tag{3.5.3}$$

and

$$\varphi_\omega(c_i - a) \geqq \varphi_\omega(c_i) < \varphi_\omega(d_i) \qquad \text{for every } i = 1, 2, \ldots, k\,,\ c_i \geqq a\,. \tag{3.5.4}$$

Thus there exist numbers t_i such that $\varphi'_\omega(t_i) = 0$, $t_i \in (c_i - a, d_i)$ for every $i = 1, 2, \ldots, k$, $c_i \geqq a$. Hence we get

$$\varphi_\omega(d_i) - \varphi_\omega(c_i) = \varphi_\omega(d_i) - \varphi_\omega(t_i) - (\varphi_\omega(c_i) - \varphi_\omega(t_i)) \leqq$$
$$\leqq \{(d_i - t_i)^2 + (c_i - t_i)^2\}\,K(\omega) \leqq 8(d_i - c_i)^2\,K(\omega):$$

$$\varphi_\omega(d_i) - \varphi_\omega(c_i) \leqq 8(d_i - c_i)^2\,K(\omega) \tag{3.5.5}$$

for every $i = 1, 2, \ldots, k$ such that $c_i \geqq a$.

The exceptional case $c_i - a < 0$ occurs only if $i = 1$, $c_1 = 0$ and is of interest for us only in the case $\tau(\omega) \in (c_1, d_1)$. However in this case there exists a $t_0 = t_0(\omega)$ (see Lemma (3.3)) such that $\varphi'_\omega(t_0) = 0$ and $|t_0 - \tau(\omega)| < a$ which implies $t_0 \in (c_1, d_1)$. Hence

$$\varphi_\omega(d_1) - \varphi_\omega(c_1) = \varphi_\omega(d_1) - \varphi_\omega(t_0) - (\varphi_\omega(c_1) - \varphi_\omega(t_0)) \leqq [(d_1 - t_0)^2 +$$
$$+ (c_1 - t_0)^2]\,K(\omega) \leqq 2(d_1 - c_1)^2$$

and thus

$$\varphi_\omega(d_1) - \varphi_\omega(c_1) \leqq 8(d_1 - c_1)^2 \text{ if } c_1 = 0,\ \tau(\omega) \in (c_1, d_1)\,. \tag{3.5.6}$$

Now if I is the set of such indices i, that $(c_i, d_i) \in B(\omega)$ and $d_i \geqq \tau(\omega)$, then from the definition of $B(\omega)$ it follows that

$$\varphi_\omega(\alpha(\omega)) - \varphi_\omega(\tau(\omega)) \leqq \sum_{i \in I} [\varphi_\omega(d_i) - \varphi_\omega(c_i)]$$

and this is according to (3.5.5) and (3.5.6) equal to or less than

$$\sum_{i \in I} 8(d_i - c_i)^2 K(\omega) .$$

However $\sum_{i \in I}(d_i - c_i)^2 \leqq [\sum_{i \in I}|d_i - c_i|]^2 \leqq [\alpha^+(\omega)]^2$ which proves (3.5.1).

It remains to prove (3.5.2). If $\tau(\omega) < +\infty$, then according to Lemma (3.3) there exists a $t_0(\omega)$ such that $\varphi'_\omega(t_0(\omega)) = 0$ and $|\tau(\omega) - t_0(\omega)| < a$. By *(3.2.2) we get two inequalities*

$$|\varphi_\omega(\tau(\omega)) - \varphi_\omega(t_0(\omega))| < a^2 K(\omega) ,$$
$$|\varphi_\omega(\beta(\omega)) - \varphi_\omega(t_0(\omega))| < (\beta(\omega) - t_0(\omega))^2 K(\omega) ,$$

which imply (3.5.2); the proof is accomplished.

(3.6) Proof of Theorem (3.1). Let $\tau = \text{Max}\,(\tau^-, \tau^+)$ and define

$$A_1 = \{\omega; \tau(\omega) > \alpha(\omega)\} \cap \Omega_0 , \qquad A_2 = \{\omega; \tau(\omega) \leqq \alpha(\omega)\} \cap \Omega_0 ,$$
$$B_{-1} = \{\omega; \tau^+(\omega) = a\} \cap \Omega_0 , \qquad B_1 = \{\omega; \tau^+(\omega) > a\} \cap \Omega_0 ,$$

We remember that (see (3.2.2) or (3.2.3))

$$\varphi_\omega(t) \leqq \varphi_\omega(0) + t\varphi'_\omega(0) + t^2 K(\omega) = \psi_\omega(t) . \tag{3.6.1}$$

If $\omega \in A_1 \cap B_{-1}$, we have $\alpha(\omega) < \tau^-(\omega)$ and even $\beta(\omega) \leqq \alpha(\omega) < \tau^-(\omega)$ by (3.1.4), which gives, according to the definition of τ^-, $\varphi_\omega(\beta(\omega)) \geqq \varphi_\omega(\alpha(\omega)$; hence by (3.6.1) we get

$$\omega \in A_1 \cap B_{-1} \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\, \varphi'_\omega(0) + \beta^2(\omega)\, K(\omega) . \tag{3.6.2}$$

If $\omega \in A_1 \cap B_1$, we have $\alpha(\omega) < \tau^+(\omega)$. Since (3.1.3) is equivalent to $\alpha(\omega) > > \beta(\omega) \Rightarrow \alpha(\omega) > \beta(\omega) \geqq \tau^+(\omega)$, we have $\alpha(\omega) \leqq \beta(\omega)$. If $\beta(\omega) \leqq \tau^+(\omega)$, then $\varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(\beta(\omega))$. If $\beta(\omega) > \tau^+(\omega)$, then $\varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(\tau^+(\omega))$ and — by (3.5.2) — $\varphi_\omega(\tau^+(\omega)) \leqq \varphi_\omega(\beta(\omega)) + 2\beta^2(\omega)\, K(\omega)$. Hence and according to (3.6.1) we get

$$\omega \in A_1 \cap B_1 \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\, \varphi'_\omega(0) + 3\beta^2(\omega)\, K(\omega) . \tag{3.6.3}$$

If $\omega \in A_2 \cap B_{-1}$, we have $\alpha(\omega) \geqq \tau^-(\omega) = \tau(\omega)$. Since $\tau(\omega)$ is finite, we may use Lemma (3.4) to get

$$\varphi_\omega(\tau^-(\omega)) \leqq \operatorname*{Min}_{t \geqq a} \psi_\omega(t) + a^2 K(\omega) \leqq \psi_\omega(\beta(\omega)) + a^2 K(\omega) \leqq$$
$$\leqq \varphi_\omega(0) + \beta(\omega)\, \varphi'_\omega(0) + 2\beta^2(\omega)\, K(\omega) .$$

Hence and according to (3.5.1) we get

$$(3.6.4) \qquad \omega \in A_2 \cap B_{-1} \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\,\varphi'_\omega(0) + \\ + \{2\beta^2(\omega) + 8[\alpha^+(\omega)]^2\}\,K(\omega)\,.$$

If $\omega \in A_2 \cap B_1$, then $\alpha(\omega) \geqq \tau^+(\omega) = \tau(\omega)$ which implies (see (3.1.3)) that also $\beta(\omega) \geqq \tau(\omega)$. Hence and according to (3.5.2) we have $\varphi_\omega(\tau(\omega)) \leqq \varphi_\omega(\beta(\omega)) + 2\beta^2(\omega)\,K(\omega)$; by (3.5.1) $\varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(\tau(\omega)) + 8[\alpha^+(\omega)]^2\,K(\omega)$ and thus

$$(3.6.5) \qquad \omega \in A_2 \cap B_1 \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\,\varphi'_\omega(0) + \\ + \{8[\alpha^+(\omega)]^2 + 2\beta^2(\omega)\}\,K(\omega)\,.$$

Finally if $\omega \in \Omega - \Omega_0$, then according to (3.1.6) $\varphi_\omega(c(\omega)) \leqq \varphi_\omega(\beta(\omega))$ and $\varphi'_\omega(c(\omega)) = 0$, which with the inequality (3.1.5) gives $\varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(c(\omega)) + (4[\alpha^+(\omega)]^2 + a^2)\,K(\omega)$. Thus $\varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(\beta(\omega)) + (4[\alpha^+(\omega)]^2 + a^2)\,K(\omega)$ and

$$(3.6.6) \qquad \omega \in \Omega - \Omega_0 \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\,\varphi'_\omega(0) + \\ + \{2\beta^2(\omega) + 4[\alpha^+(\omega)]^2\}\,K(\omega)\,.$$

Since $(A_1 \cap B_{-1}) \cup (A_1 \cap B_1) \cup (A_2 \cap B_{-1}) \cup (A_2 \cap B_1) \cup (\Omega - \Omega_0) = \Omega$, the relations (3.6.2) to (3.6.6) give

$$(3.6.7) \qquad \omega \in \Omega \Rightarrow \varphi_\omega(\alpha(\omega)) \leqq \varphi_\omega(0) + \beta(\omega)\,\varphi'_\omega(0) + \{8[\alpha^+(\omega)]^2 + \\ + 3\beta^2(\omega)\}\,K(\omega)$$

Hence

$$(3.6.8) \qquad f(X(\omega) + \alpha(\omega)\,Y(\omega)) \leqq f(X(\omega)) + \beta(\omega)\,\langle Y(\omega), D(X(\omega))\rangle + \\ + K(8[\alpha^+(\omega)]^2 + 3\beta^2(\omega))\,\|Y(\omega)\|^2$$

and by (3.1.1) and (3.1.2)

$$(3.6.9) \qquad \mathbf{E}_{\mathscr{X}_n, Y_n} f(X + \alpha Y) \leqq f(X) + a_n\,\langle \mathbf{M}_n(\mathscr{X}_n), D(X)\rangle + 11 c_n K\,\|Y\|^2\,,$$

which implies (3.1.7). The theorem is proved.

4. Particular choices of length α_n of the n-th step

(4.1) **Theorem.** *Suppose f satisfies Assumption* (2.1) *and $\alpha_n = a_n$ is a number. Then* (3.1.7) *holds with* $c_n = \dfrac{a_n^2}{11}$, i. e.

$$(4.1.1) \qquad \mathbf{E}_{\mathscr{X}_n} f(X_{n+1}) \leqq f(X_n) + a_n\,\langle \mathbf{M}_n(\mathscr{X}_n), D(X_n)\rangle + a_n^2 \mathbf{E}_{\mathscr{X}_n}\,\|Y_n\|^2\,.$$

Proof. Since $X_{n+1} = X_n + a_n Y_n$, (4.4.1) follows directly from (2.1.1).

In the preceding theorem a simple way of choosing α_n is described, which was hithertho used by authors proposing approximation schemes. However a more refined definition of α_n can save us observations, especially, if the number q of dimensions is large and if X_n is far from the extremal point we seek.

In the following theorem we describe such a method. We note that the condition (4.2.1) will be satisfied for example if $m_i = f(X_n + iaY_n)$. The random variables V_i can be called estimates of m_i, or especially of $f(X_n + iaY_n)$. The generality obtained by introducing the variables m_i is useful in the cases in which f cannot be observed and observations of another function R, related to f, are at our disposition.

(4.2) **Theorem.** *Let f satisfy Assumption* (2.1), *let n be a natural number, a, d, c real positive numbers. Let V_i, $m_i (i = 1, 2, \ldots)$ be random variables such that for every $j = 1, 2, 3, \ldots; i = -1, 1; i + j \geqq 1$*

(4.2.1)
$$f(X_n(\omega) + jaY_n(\omega)) > f(X_n(\omega) + (j + i)\, aY_n(\omega)) \Rightarrow \Rightarrow m_j(\omega) \geqq m_{j+i}(\omega)$$

and that

(4.2.2)
$$\sum_{k=1}^{\infty} k^s P_{\mathscr{X}_n, Y_n} \{V_1 - m_1 > V_2 - m_2 > \ldots > V_k - m_k \leqq \leqq V_{k+1} - m_{k+1}\} = 1 \text{ for } s = 0, \ = d \text{ for } s = 1, \leqq c \text{ for } s = 2 .$$

Define $\alpha_n(\omega) = ka$ for ω in the set

(4.2.3)
$$A_k = \{V_1 > V_2 > \ldots > V_k \leqq V_{k+1}\}$$

and suppose that

(4.2.4)
$$\mathbf{E}_{\mathscr{X}_n, Y_n}[\alpha_n^+]^2 \leqq ca^2 ,$$

Then (3.1.7) *holds for $a_n = ad$ and $c_n = a^2 c$.*

Proof. Define $\beta(\omega) = ka_n$ if ω is in the set

(4.2.5) $B_k = \{V_1 - m_1 > V_2 - m_2 > \ldots > V_k - m_k \leqq V_{k+1} - m_{k+1}\}$.

For the proof of the theorem it suffices to show that assumptions of Theorem (3.1) are satisfied. The condition (3.1.1) is repeated in (4.2.4), (3.1.2) follows from (4.2.2), f satisfies Assumption (2.1) and it remains to prove that (3.1.3) to (3.1.6) hold for some $\Omega_0 \subset \Omega$: we shall show it for $\Omega_0 = \Omega$, in which case (3.1.5) and (3.1.6) are trivial.

We shall prove (3.1.3); let $\omega \in \Omega$, $ja = \beta(\omega) < \tau^+(\omega)$. Then $f(X_n(\omega) + \beta(\omega) \,.\, Y_n(\omega)) < f(X_n(\omega) + (\beta(\omega) + a)\, Y_n(\omega))$ and according to (4.2.1)

(4.2.6)
$$m_j(\omega) \leqq m_{j+1}(\omega) .$$

From the definition of β it follows that $V_j(\omega) - m_j(\omega) \leqq V_{j+1}(\omega) - m_{j+1}(\omega)$ and according to (4.2.6) $V_j(\omega) \leqq V_{j+1}(\omega)$. Thus $\alpha_n(\omega) \leqq ja = \beta(\omega)$ and (3.1.3) is proved.

It remains to prove (3.1.4). Let $\omega \in \Omega$, $ja = \alpha_n(\omega) < \tau^-(\omega)$. Then $f(X_n(\omega) + \alpha_n(\omega)\, Y_n(\omega)) > f(X_n(\omega) + (\alpha_n(\omega) + 1)\, Y_n(\omega))$ and — according to (4.2.1) — $m_j(\omega) \geqq m_{j+1}(\omega)$, which with the obvious inequality $V_j(\omega) \leqq V_{j+1}(\omega)$ gives

$V_j(\omega) - m_j(\omega) \leqq V_{j+1}(\omega) - m_{j+1}(\omega)$. But the last inequality implies $\beta(\omega) \leqq \leqq ja = \alpha_n(\omega)$ and the proof of (3.1.4) and of the whole theorem is accomplished.

The preceding theorem imposes some very weak conditions on the estimates V_i of m_i. Their generality will be apparent in the next theorem.

(4.3) **Theorem.** *Let f, m_i satisfy the conditions of the preceding theorem, let V_i be random variables such that $\tilde{V}_i = V_i - m_i$ are distributed independently, identically and continuously and are independent of $\mathscr{X}_n$, Y_n. Then* (4.2.2) *holds with*

$$d = \sum_{k=1}^{\infty} k^2 \frac{1}{(k+1)!}, \quad c = \sum_{k=1}^{\infty} \frac{k^2}{2^{k-1}} \tag{4.3.1}$$

If α_n is defined as in the preceding theorem, (4.2.4) *and* (3.1.7) *hold with $a_n = ad$ and $c_n = a^2 c$.*

Proof. Obviously

$$P_{\mathscr{X}_n, Y_n}(\tilde{V}_1 > \tilde{V}_2 > \ldots > \tilde{V}_k \leqq \tilde{V}_{k+1}) = P(\tilde{V}_1 > \tilde{V}_2 > \ldots > \tilde{V}_k \leqq \tilde{V}_{k+1}) = = \frac{k}{(k+1)!},$$

which implies (4.2.2) with c and d given by (4.3.1). We shall show that (4.2.4) holds. Denote $\varphi_j = f(X_n + jaY_n)$ and let

$$\bigcup_{i=1}^{K(\omega)} \langle n_i(\omega), n_i(\omega) + h_i(\omega)) = \bigcup \{\langle j-1, j);\ j = 2, 3, \ldots, \varphi_{j-1}(\omega) < \varphi_j(\omega)\} \tag{4.3.2}$$

where n_i, h_i are natural number valued functions on Ω, $K(\omega)$ is a natural number or $+\infty$ and

$$n_1 < n_1 + h_1 < n_2 < \ldots. \tag{4.3.3}$$

It is easy to see that, by the relations (4.3.2) and (4.3.3), K, n_i and h_i are uniquely determined random variables (K possibly infinite) and that they are functions of $[\mathscr{X}_n, Y_n]$ only. Further we have

$$a^{-1}\alpha_n^+ = \gamma + \sum\left\{h_i;\ i = 1, 2, \ldots, n_i + h_i \leqq \frac{\alpha_n}{a}\right\}, \tag{4.3.4}$$

where γ is positive part of $\varphi_1 - \varphi_0$:

Thus, denoting by N the index for which $\sum_{i=1}^{N} h_i = k - \gamma$, the event $\{\alpha_n^+ = ak\}$ implies the event

$$\bigcap_{j=1}^{N} \left\{n_j + h_j < \frac{\alpha_n}{a}\right\} \tag{4.3.5}$$

and this implies, according to the definition of α_n, the event

$$\bigcap_{j=1}^{N} \{V_{n_j} > V_{n_j+1} > \ldots > V_{n_j+h_j}\} \tag{4.3.6}$$

However from (4.3.2) it follows that

(4.3.7) $$\varphi_{n_j} < \varphi_{n_j+1} < \ldots < \varphi_{n_j+h_j}$$

which gives according to (4.2.1) the inequality

4.3.8) $$m_{n_j} \leqq m_{n_j+1} \leqq \ldots \leqq m_{n_j+h_j}.$$

Since $\tilde{V}_i = V_i - m_i$, the event in (4.3.6) implies the following

(4.3.8) $$\bigcap_{j=1}^{N} \{\tilde{V}_{n_j} > \tilde{V}_{n_j+1} > \ldots > \widehat{V}_{n_j+h_j}\}$$

and we get that

(4.3.9) $$P_{\mathscr{X}_n, Y_n} \{\alpha_n^+ = ak\} \leqq P_{\mathscr{X}_n, Y_n} \dot{\bigcap_{j=1}^{N}} \{\tilde{V} \; > \tilde{V}_{n_j+1} > \ldots$$
$$\ldots > \tilde{V}_{n_j+h_j}\} = \prod_{j=1}^{N(\omega)} \frac{1}{(h_j(\omega)+1)!},$$

the last equality being due to the fact that $\tilde{V}_{n_j}$ are independent, continuous and that the sequence $\tilde{V}_1, \tilde{V}_2, \ldots$, is independent of $[\mathscr{X}_n, Y_n]$, n_i, h_i, N. Thus $P_{\mathscr{X}_n, Y_n} \{\alpha_n^+ = ak\}$ has an upper bound of

$$\frac{1}{(h_1(\omega)+1)!\,(h_2(\omega)+1)!\ldots(h_{N(\omega)}(\omega)+1)!}$$

where $\sum_{i=1}^{N(\omega)} h_i(\omega) \geqq k - 1$. Thus there are at least $k - 1$ factors greater than 2 in the denominator, which implies

(4.3.10) $$P_{\mathscr{X}_n, Y_n} \{\alpha_n^+ = ak\} \leqq \frac{1}{2^{k-1}},$$

whence (4.2.4) follows with c defined by (4.3.1). Since the last assertion of the Theorem follows from Theorem (4.2), the proof is accomplished.

The two theorems already proved deal with the problem of approximating a point at which the function estimated acquires its minimum. An analogous result for the situation of the Robbins — Monro procedure is given in the following theorem:

(4.4) **Theorem.** *Let f be a function defined on E_1 staisfying Assumption* (2.1), *decreasing in* $(-\infty, \Theta)$ *and increasing in* $(\Theta, +\infty)$. *Let n be a natural number, a a positive number and V_i, m_i $(i = 1, 2, \ldots)$ random variables such that* $\widehat{V}_i = \operatorname{sign}(V_i - m_i)$ *are independently and identically distributed random variables independent of $\mathscr{X}_n$, Y_n with* $\mathsf{E}\widehat{V}_i = 0$. *Suppose that if f is increasing resp. decreasing in the point $X_n(\omega) + jaY_n(\omega)$, then $m_j(\omega)$ is non-negative resp. non-positive.*

Let $\alpha_n(\omega) = ja$ *for* ω *such that*

$$-\operatorname{sign} Y_n(\omega) = \operatorname{sign} V_1(\omega) = \ldots = \operatorname{sign} V_{j-1}(\omega) \neq \operatorname{sign} V_j(\omega)$$

(*if* $-\operatorname{sign} Y_n(\omega) \neq V_1(\omega)$, *we put* $\alpha_n(\omega) = 1$). *Then* (3.1.7) *holds with* $a_n = 2a$, $c_n = 6a^2$.

Proof. Let us denote by Ω_0 the set of those $\omega \in \Omega$, for which the interval $(X_n(\omega), X_n(\omega) + \alpha_n(\omega) Y_n(\omega)\rangle \cup \langle X_n(\omega) + \alpha_n(\omega) Y_n(\omega), X_n(\omega))$ is non-empty (i. e. $Y_n(\omega) \neq 0$) and does not contain Θ. Further put $\beta(\omega) = ja$ for such ω that

$$-\operatorname{sign} Y_n(\omega) = \widehat{V}_1(\omega) = \ldots = \widehat{V}_{j-1}(\omega) \neq \widehat{V}_j(\omega) .$$

We shall prove that for our Ω_0, β the relations (3.1.3) and (3.1.4) hold. If $\omega \in \Omega_0$ and $ja = \beta(\omega) < \tau^+(\omega)$, then $\tau^+(\omega) > a$, i. e. either (i) $X_n(\omega) \leqq \Theta$ and $\operatorname{sign} Y_n(\omega) = -1$ or (ii) $X_n(\omega) \geqq \Theta$ and $\operatorname{sign} Y_n(\omega) = 1$. In the case (i) $1 \neq \widehat{V}_j(\omega)$, i. e. $V_j(\omega) \leqq m_j(\omega)$ and since $X_n(\omega) + \beta(\omega) Y_n(\omega) < \Theta$ and thus $m_j(\omega) \leqq 0$, we get $V_j(\omega) \leqq 0$, $\operatorname{sign} V_j(\omega) \neq -\operatorname{sign} Y_n(\omega)$, which implies that $\alpha_n(\omega) \leqq aj = \beta_n(\omega)$. Thus (3.1.3) is proved in the case (i). The proof in the case (ii) is analogous and will be omitted. Now turn to (3.1.4). If $\omega \in \Omega_0$ and $aj = \alpha_n(\omega) < \tau^-(\omega)$ we have either (i) $X_n(\omega) < \Theta$, $Y_n(\omega) > 0$, $X_n(\omega) + \alpha_n(\omega) Y_n(\omega) < \Theta$ or (ii) $X_n(\omega) \geqq \Theta$, $Y_n(\omega) < 0$, $X_n(\omega) + \alpha_n(\omega) Y_n(\omega) > \Theta$. In the case (i) $-1 \neq \operatorname{sign} V_j(\omega)$, i. e. $V_j(\omega) \geqq 0$ and since $m_j(\omega) \leqq 0$, we have $\widehat{V}_j(\omega) > -1$ so that $\beta(\omega) \leqq ja_n = \alpha_n(\omega)$. (3.1.4) is proved in the case (i); the proof for the case (ii) is similar and is omitted.

Now if $\omega \in \Omega - \Omega_0$, then either $Y_n(\omega) = 0$ and in this case (3.1.5) and (3.1.6) are satisfied by taking $c(\omega) = \alpha_n(\omega)$, or the interval $(X_n(\omega), X_n(\omega) + \alpha_n(\omega) \,.\, Y_n(\omega)\rangle \cup \langle X_n(\omega) + \alpha_n(\omega) Y_n(\omega), X_n(\omega))$ contains Θ. In the last case (3.1.5) and (3.1.6) are satisfied by $c(\omega) = \dfrac{\Theta - X_n(\omega)}{Y_n(\omega)}$.

Finally it is easy to see that

$$\mathbf{E}_{\mathscr{X}_n, Y_n}\beta = \mathbf{E}\beta = a \sum_{j=1}^{\infty} j \left(\frac{1}{2}\right)^j = 2a = a_n ,$$

$$\mathbf{E}_{\mathscr{X}_n, Y_n}\beta^2 = \mathbf{E}\beta^2 = a^2 \sum_{j=1}^{\infty} j^2 \left(\frac{1}{2}\right)^j = 6a^2 = c_n$$

and

$$\mathbf{E}_{\mathscr{X}_n, Y_n}[\alpha_n^+]^2 \leqq \mathbf{E}\beta^2 = c_n .$$

Since f satisfies Assumption (2.1), all conditions of Theorem (3.1) hold and (3.1.7) is proved.

We have seen that, if f satisfies Assumption (2.1) and if for every n α_n are chosen in one of the ways described in Theorems (4.1) to (4.4) (not necessarily in a unique way for every n), then the following assumption holds.

(4.5) **Assumption.** For every n, the relation

$$\mathsf{E}_{\mathscr{X}_n} f(X_{n+1}) \leqq f(X_n) + a_n \langle \mathbf{M}_n(\mathscr{X}_n), D(X_n) \rangle + a_n^2 \, C \mathsf{E}_{\mathscr{X}_n} \|Y_n\|^2 ,$$

holds, where a_n, C are positive numbers.

5. Convergence theorems. The following lemma and theorem are slight modifications of Blum's [2] results.

(5.1) **Lemma.** *Let ξ_n be non-negative random variables and let*

(5.1.1) $$\sum_{n=1}^{\infty} \mathsf{E}\Theta_n^+ < +\infty ,$$

where Θ_n^+ denotes the non-negative part of the random variable

(5.1.2) $$\Theta_n = \mathsf{E}_{\xi_1, \xi_2, \ldots, \xi_{n-1}} \xi_n - \xi_{n-1} .$$

Then the sequence ξ_i converges to a random variable ξ.

Proof. Put $\vartheta_i = \sum_{j=1}^{i} \Theta_j^+$ and $\zeta_i = \vartheta_i - \xi_i$. We have $\zeta_n = \vartheta_n - \xi_n = \vartheta_{n-1} + \Theta_n^+ - \xi_n \geqq \vartheta_{n-1} + \Theta_n - \xi_n = \vartheta_{n-1} + \Theta_n - \xi_{n-1} - (\xi_n - \xi_{n-1}) = \zeta_{n-1} + \Theta_n - (\xi_n - \xi_{n-1})$. According to (5.1.2) we have $\mathsf{E}_{\xi_1, \ldots, \xi_{n-1}} [\Theta_n - (\xi_n - \xi_{n-1})] = 0$ and thus

$$\mathsf{E}_{\xi_1, \xi_2, \ldots, \xi_{n-1}} \zeta_n \geqq \mathsf{E}_{\xi_1, \xi_2, \ldots, \xi_{n-1}} (\zeta_{n-1} + \Theta_n - (\xi_n - \xi_{n-1})) = \mathsf{E}_{\xi_1, \xi_2, \ldots, \xi_{n-1}} \zeta_{n-1} ;$$

however $\zeta_1, \ldots, \zeta_{n-1}$ are functions of $\xi_1, \ldots, \xi_{n-1}$ only and thus

(5.1.3) $$\mathsf{E}_{\zeta_1, \zeta_2, \ldots, \zeta_{n-1}} \zeta_n \geqq \zeta_{n-1} ,$$

which shows that the sequence $\zeta_1, \zeta_2, \ldots$, is a semimartingale. Now (5.1.1) guarantees that $\sup_n \mathsf{E}\vartheta_n < +\infty$ which implies, since ξ_n are non-negative, that $\sup_n \mathsf{E}\zeta_n^+ < +\infty$. On the other hand

$$\mathsf{E}\,(|\zeta_n| - \zeta_n^+) \leqq \mathsf{E}\xi_n = \mathsf{E}\vartheta_n - \mathsf{E}\zeta_n \leqq \sup_n \mathsf{E}\vartheta_n - \mathsf{E}\zeta_1 ,$$

for by (5.1.3) $\mathsf{E}\zeta_1 \leqq \mathsf{E}\zeta_2 \leqq \ldots$. But in this way we have proved that

(5.1.4) $$\sup_n \mathsf{E}\,|\zeta_n| < +\infty , \quad \sup_n \mathsf{E}\xi_n < +\infty .$$

From the first of the inequalities it follows by the martingale theorem (see Theorem 4.1, Assertion I of Doob [6]), that ζ_n is a convergent sequence. Since the non-decreasing sequence ϑ_n converges according to (5.1.1), ξ_n also converges to a random variable.

(5.2) **Theorem.** *Let Assumptions* (2.1) *and* (4.5) *hold, let B_n be non-negative functions on Ω, let b_n, d_n, e_n, K_2 be positive numbers and let*

(5.2.1) $$\langle \mathbf{M}_n(\mathscr{X}_n), D(X_n) \rangle \leqq -B_n^2 + b_n(K_2 + B_n) ,$$

(5.2.2) $$\|\mathbf{M}_n(\mathscr{X}_n)\|^2 \leqq d_n + e_n B_n^2 ,$$

$$(5.2.3) \qquad \mathbf{E}_{\mathscr{X}_n} \|Y_n - M(\mathscr{X}_n)\|^2 \leqq d_n + e_n B_n^2 ,$$

$$(5.2.4) \qquad \Sigma a_n = +\infty, \ \Sigma a_n b_n < +\infty, \ \Sigma a_n^2 d_n < +\infty, \ \lim b_n = \lim a_n e_n = 0 .$$

Then there exists a sequence n_i *and a set* $\Omega_0 \subset \Omega$ *such that* $P(\Omega_0) = 1$ *and that* $\lim_{n\to\infty} f(X_n(\omega))$ *exists and is finite and* $\lim_{i\to\infty} B_{n_i}(\omega) = 0$ *for every* $\omega \in \Omega_0$.

If the functions B_n *depend on* X_n *only, i. e. if* $B_n = B(X_n)$, *where* B *is a function on* $\mathbf{X}$, *then for every* $\omega \in \Omega_0$

$$(5.2.5) \qquad \lim f(X_n) \in \{a;\ x_i \in \mathbf{X},\ x_i \to x \in \mathbf{X},\ a = f(x),\ B(x_i) \to 0\} \cup F ,$$

where $F = \{a;\ x_i \in \mathbf{X},\ \|x_i\| \to +\infty,\ f(x_i) \to a \in E_1,\ B(x_i) \to 0\}$.

If B *is continuous, then*

$$(5.2.6) \qquad \lim f(X_n(\omega)) \in f(\{x;\ B(x) = 0\}) \cup F .$$

Proof. By Assumption (4.5) we have

$$\mathbf{E}_{\mathscr{X}_n} f(X_{n+1}) \leqq f(X_n) + a_n \langle \mathbf{M}_n(\mathscr{X}_n), \quad D(X_n)\rangle + a_n^2 C \mathbf{E}_{\mathscr{X}_n} \|Y_n\|^2 .$$

From (5.2.2) and (5.2.3) we get

$$\mathbf{E}_{\mathscr{X}_n} \|Y_n\|^2 \leqq 2d_n + 2e_n B_n^2$$

and thus

$$\mathbf{E}_{\mathscr{X}_n} f(X_{n+1}) \leqq f(X_n) + a_n(-B_n^2 + b_n(K_2 + B_n)) + 2a_n^2(d_n + e_n B_n^2)\, C = \\ = -a_n(1 - 2a_n e_n C) . \left(B_n^2 - \frac{b_n}{1 - 2a_n e_n C} B_n - \frac{2a_n d_n C + b_n K_2}{1 - 2a_n e_n C}\right)$$

(since only limiting properties are of interest, we may assume with respect to (5.2.4) that $2a_n e_n C < 1$). Putting

$$\mu_n = a_n(1 - 2a_n e_n C) ,$$

$$\nu_n = \frac{b_n}{1 - 2a_n e_n C},$$

$$\varrho_n = \frac{2a_n \mathrm{d}_n C + b_n K_2}{1 - 2a_n e_n C},$$

we get

$$(5.2.7) \qquad \mathbf{E}_{\mathscr{X}_n}(f(X_{n+1}) - f(X_n)) \leqq -\mu_n[B_n^2 - \nu_n B_n - \varrho_n] ,$$

where μ_n, ν_n, ϱ_n are positive numbers, satisfying according to (5.2.4) the relations

$$(5.2.8) \qquad \Sigma\mu_n = +\infty, \ \Sigma\mu_n\nu_n < +\infty, \ \Sigma\mu_n\varrho_n < +\infty, \ \lim \nu_n = 0 .$$

Since only limiting properties are of interest, we may assume that $\nu_n < \frac{1}{2}$ for all n. Put

$$\lambda_n = \begin{cases} 1 & \text{if } B_n > 1 , \\ 0 & \text{if } B_n \leqq 1 ; \end{cases}$$

then

$$(5.2.9) \qquad (1-\lambda_n)\,\nu_n B_n \leqq \nu_n,\ B_n^2 - \lambda_n\nu_n B_n \geqq \frac{1}{2}B_n^2 .$$

Since according to (5.2.7) $\mathbf{E}_{\mathscr{X}_n}(f(X_{n+1}) - f(X_n)) \leqq -\mu_n[B_n^2 - \lambda_n\nu_n B_n - \varrho_n] + \mu_n(1-\lambda_n)\,\nu_n B_n$, we get by (5.2.9)

$$(5.2.10) \qquad \mathbf{E}_{\mathscr{X}_n}(f(X_{n+1}) - f(X_n)) \leqq -\frac{1}{2}\mu_n B_n^2 + \mu_n(\nu_n + \varrho_n) .$$

Hence we get, since $\mu_n(\nu_n + \varrho_n) > 0$,

$$(5.2.11) \qquad \{\mathbf{E}_{f(X_1),f(X_2),\ldots,f(X_n)}(f(X_{n+1}) - f(X_n))\}^+ \leqq \mu_n(\nu_n + \varrho_n) ,$$

where on the right we have a summable sequence $\mu_n(\nu_n + \varrho_n)$. This is (see Lemma (5.1)) a sufficient condition for the sequence $f(X_n)$ to be convergent.

Now let us denote

$$(5.2.12) \qquad C_n = \frac{-\mathbf{E}_{\mathscr{X}_n}f(X_{n+1}) + f(X_n) + \mu_n(\nu_n + \varrho_n)}{\mu_n} .$$

By 5.2.10 we have $0 \leqq \frac{1}{2}B_n^2 \leqq C_n$ and

$$\mathbf{E}_{\mathscr{X}_n}f(X_{n+1}) - f(X_n) \leqq -\mu_n C_n + \mu_n(\nu_n + \varrho_n) ,$$
$$\mathbf{E}(f(X_{n+1}) - f(X_n)) \leqq -\mu_n\mathbf{E}C_n + \mu_n(\nu_n + \varrho_n) ,$$
$$\mathbf{E}f(X_{n+1}) \leqq f(X_1) - \sum_{j=1}^{n}\mu_n\mathbf{E}C_n + \sum_{j=1}^{n}\mu_j(\nu_j + \varrho_j) .$$

Since by (5.2.8) $0 < \sum_{j=1}^{\infty}\mu_j(\nu_j + \varrho_j) < +\infty$ and since $f(X_{n+1}) \geqq 0$, the non-positive term $-\sum_{j=1}^{n}\mu_j\mathbf{E}C_j$ converges, too, which implies the existence of a sequence m_j such that $\mathbf{E}C_{m_j} \to 0$, whence it follows that there exists a $\Omega_1 \subset \Omega$ with $P(\Omega_1) = 1$ and a sequence n_i such that $C_{n_i}(\omega) \to 0$ for every $\omega \in \Omega_1$. However the inequality $B_n^2 \leqq 2C_n$ implies that $B_{n_i}(\omega) \to 0$ for every $\omega \in \Omega_1$. Formerly we have proved that there exists a $\Omega_2 \subset \Omega$ of probability one and such that $f(X_n(\omega))$ converges to a number if $\omega \in \Omega_2$. Clearly $f(X_n(\omega))$ converges to a number and $B_{n_i}(\omega) \to 0$ for every $\omega \in \Omega_0 = \Omega_1 \cap \Omega_2$ and $P(\Omega_0) = 1$.

Finally let $B_n = B(X_n)$ for a function B and let $\omega \in \Omega_0$. We may choose a subsequence x_i of the sequence $X_{n_i}(\omega)$ such that either $x_i \to x \in \mathbf{X}$ or $\|x_i\| \to +\infty$. In the first case we get by the continuity of f that $\lim f(X_n(\omega)) = \lim f(x_i) = f(x)$ and $\lim B(x_i) = 0$. Hence the relation (5.2.5) follows. Since (5.2.6) is a direct consequence of (5.2.5) and of the assumed continuity of B, the proof is accomplished.

(5.3) Note. In the preceding Theorem (5.2.1) with $b_n \to 0$ is a basic condition, which ensures that $\langle \mathbf{M}_n(\mathscr{X}_n), D(X_n)\rangle$ is negative at least if n and B_n are large.

Hence, from (5.2.7) and from the non-negativity of f it was then possible to deduce that for every ω there exists a sequence n_i such that $B_{n_i}(\omega) \to 0$. This is of interest for example if $\mathbf{M}_n(\mathscr{X}_n)$ is such that $B_n = \|D(X_n)\|$. In this case (and if certain conditions are satisfied, in a more general case $B_n = \|H_n(\mathscr{X}_n)\, D(X_n)\|$ (see also the following theorem) the condition (5.2.6) can be written as

$$\lim_{n\to\infty} f(X_n(\omega)) \in A \cup A_1 , \tag{5.3.1}$$

where

$$A = f(\{x; D(x) = 0\}) \tag{5.3.2}$$

and

$$A_1 = \{a; x_i \in \mathbf{X}, \|x_i\| \to +\infty, D(x_i) \to 0, f(x_i) \to a \in E_1\} . \tag{5.3.3}$$

It is easy to see that if $\|X_n(\omega)\|$ is bounded for every $\omega \in \Omega_0$ (and this condition will be satisfied if e. g. the assumptions of Theorem (5.5) hold), then (5.3.1) can be strengthened to

$$\lim_{n\to\infty} f(X_n(\omega)) \in A = f(\{x; D(x) = 0\}) . \tag{5.3.4}$$

However in certain cases (5.3.4) can be also deduced from (5.3.2) and (5.3.3). For example, in Sections (6.1), (6.2), (7.1) and (7.2), $\|x_i\| \to \infty$ implies $f(x_i) \to +\infty$ so that $A_1 = \emptyset$. Similarly if $\inf \{\|D(x)\|; \|x - \Theta\| > \varepsilon\} > 0$ for every $\varepsilon > 0$ (see (1.4.1) for $R = f$), then again $A_1 = \emptyset$ and (5.3.4) holds; moreover

$$\lim_{n\to\infty} f(X_n(\omega)) = f(\Theta) \tag{5.3.5}$$

and f has at Θ its absolute minimum. If further conditions are satisfied, e. g. that $\inf \{|f(\Theta) - f(x)|; \|x - \Theta\| > \varepsilon\} > 0$ for every $\varepsilon > 0$ (see (1.4.1)), then (5.3.5) implies

$$X_n(\omega) \to \Theta . \tag{5.3.6}$$

In this connection we remark that, from the practical point of view, usually not the distance between X_n and Θ is of interest, but the distance between $f(X_n)$ and $\inf_{x\in\mathbf{X}} f(x)$. The second formulation avoids certains unessential difficulties which arise with the first. For example if $f(y) = \inf_{x\in\mathbf{X}} f(x)$ for every y in a convex set containing more than one point, we are not able to establish a relation of the form (5.3.6), although from the practical point of view such a situation may be considered as agreable for stability reasons.

Let us return to the relation (5.3.4). The case in which the set $\{x; Dx = 0\}$ consists of a single point Θ (the non-negativity of f then implies that if f aquires its absolute minimum it does so at Θ) has been discussed. However, there are many practical situations in which $\{x; Dx = 0\}$ consists of more than one point .In this case, it is natural that in general the relation $\lim_{n\to\infty} f(X_n) = \inf_{x \in \mathbf{X}} f(x)$ does not hold and that $\lim f(X_n(\omega))$ may converge to $f(x)$, where at x f acquires its local minimum. Indeed in such a situation there is, in our opinion, no other way to approximate the point of the absolute minimum than a systematical estimation of $f(x)$ for every x in a reasonably dense net in $\mathbf{X}$. Let us denote by A_+ and A_- the sets of points at which f has its local maximum and minimum respectively. As we mentioned, we have no chance to prove that $\lim f(X_n) = \inf_{x \in \mathbf{X}} f(x)$. By Theorem (5.5) it is easy to construct examples showing that every effort to prove that $\lim f(X_n(\omega)) \in f(A_- - A_+)$ would also be unsuccessful. However we did not even succeed in proving $\lim f(X_n(\omega)) \in f(A_-)$ for almost all $\omega \in \Omega$, which is perhaps a consequence of the fact, that the method of proving Theorem (5.2) is based on the first derivative D, which does not distinguish between the points of A_+ and A_-.

The next theorem will sometimes be useful in verifying the conditions of Theorem (5.2).

(5.4) **Theorem.** *Let n be a natural number and let for every $x \in \mathbf{X}^n$ $H_n^2(x)$ be a non-negative hermitian matrix, i. e. let $\langle H_n^2(x)a, b\rangle = \langle H_n(x)\, a, H_n(x)\, b\rangle$ for every $x \in \mathbf{X}^n$, $a, b \in \mathbf{X}$. Let*

$$\mathbf{M}_n(\mathscr{X}_n) = - H_n^2(\mathscr{X}_n)\, D(X_n) + h_n \Theta_n(\mathscr{X}_n)\,, \tag{5.4.1}$$

where Θ_n is a matrix function on $\mathbf{X}^n$, h_n is a number and

$$\|\Theta_n(\mathscr{X}_n)\| \leqq 1\,, \quad h_n \geqq 0\,. \tag{5.4.2}$$

Let further

$$\|D(X_n)\| \leqq g_n(C_1 + \|H_n(\mathscr{X}_n)\, D(X_n)\|)\,, \tag{5.4.3}$$

where g_n, C_1 are non-negative numbers. Then (5.2.1) *holds with*

$$B_n = \|H_n(\mathscr{X}_n)\, D(X_n)\|\,, \quad K_2 = C_1, \quad b_n = h_n g_n\,. \tag{5.4.4}$$

Proof. If (5.4.3) holds, then from (5.4.1) we get

$$\begin{aligned}\langle \mathbf{M}_n(\mathscr{X}_n), D(X_n)\rangle &= \langle - H_n^2(\mathscr{X}_n)\, D(X_n), D(X_n)\rangle + h_n \langle \Theta_n(\mathscr{X}_n), D(X_n)\rangle \leqq \\ &\leqq - \langle H_n(\mathscr{X}_n)\, D(X_n),\ H_n(\mathscr{X}_n)\, D(X_n)\rangle + h_n \|D(X_n)\| \leqq \\ &\leqq - \|H_n(\mathscr{X}_n)\, D(X_n)\|^2 + h_n g_n (C_1 + \|H_n(\mathscr{X}_n)\, D(X_n)\|)\,,\end{aligned}$$

so that (5.2.1) holds with B_n, b_n and K_2 defined by (5.4.4).

(5.5) **Theorem.** *Let $\mathbf{M}_n(\mathscr{X}_n) = \mathbf{N}_n(X_n)$, where $\mathbf{N}_n$ are function on $\mathbf{X}$, satisfy* (5.2.2) *and* (5.2.3) *with $e_n = 0$, let* (5.2.4) *hold. Let there exist a $r > 0$ such that for every $i = 1, 2, \ldots, q$ $\operatorname{sign} \mathbf{N}^{(i)}(x) \,.\, \operatorname{sign} x^{(i)} \leqq 0$ if $|x^{(i)}| > r$. Let $\alpha_n = a_n$, or let α_n be defined as in Theorem* (4.3) *with $a = \frac{\alpha_n}{d}$ and with such m_i*

that (instead of (4.2.1)) if denoting by $\delta(x)$ *the distance of* x *from the set* $\{x; x \in \mathbf{X}, |x^{(i)}| < r\}$ *we have for every* $i = -1, 1; j = 1, 2, \ldots; i + j \geqq 1; \omega \in \Omega$

$$(5.5.1) \qquad \delta(X_n(\omega) + jaY_n(\omega)) > \delta(X_n(\omega) + (j + i)\, Y_n(\omega)) \Rightarrow m_j(\omega) \geqq \geqq m_{j+1}(\omega)\,.$$

Then there exists a subset $\Omega_0 \subset \Omega$ *of probability one and such that* $\sup_{n=1,2,\ldots} \|X_n(\omega)\| < + \infty$ *for every* $\omega \in \Omega_0$.

Proof. Put $f(x) = \delta^2(x) = \sum_{i=1}^{q} (|x^{(i)}| - r)_+^2$.[5] Clearly f satisfies Assumption (2.1) and

$$D^{(i)}(x) = 2(|x^{(i)}| - r)_+ \operatorname{sign} x^{(i)}\,,$$

so that (5.2.1) is satisfied with $B_n = 0$, $b_n = 0$. From Theorem (4.1) or (4.3) it follows that Assumption (4.5) holds. We may apply Theorem (5.2) and the boundedness of $\|X_n(\omega)\|$ for almost every ω follows from the convergence of $f(X_n)$.

The simple condition concerning sign $\mathbf{N}^{(i)}(x)$ is satisfied e. g. in the case of the search for a minimum of a function R, if sign $\mathbf{N}_n^{(i)}(X_n) = \operatorname{sign}\, [R(X_n) - R(X_n + c_n\Delta^{(i)})]$ and if sign $D^{(i)}R(x)\,.\operatorname{sign} x^{(i)} \geqq 0$ for $|x^{(i)}| > \frac{r}{2} > c_n$. In this case also (5.5.1) is satisfied if $a < \frac{r}{2}$, $m_j = R(X_n + jaY_n)$.

6. The Robbins-Monro method and its modifications. (6.1) Suppose that R, Y_n, a_n satisfy the conditions of Theorem (1.2), but let us require

$$(6.1.1) \qquad R(x) \leqq 0 \text{ for } x < \Theta\,, \quad R(x) \geqq 0 \text{ for } x > 0$$

instead of the stronger condition (1.2.1). Define $f(x) = (x - \Theta)^2$. Then minimizing f is formally equivalent to solving the equation $R(x) = 0$.

Suppose further that α_n are chosen in such a way that Assumption (4.5) holds with a suitable constant C. Theorem (4.1) says that this is so if $\alpha_n = a_n$ as in Theorem (1.2). However this is not the unique possible choice of α_n as we have proved in Theorem (4.5). Thus we may, after determining the value $Y_n(\omega)$, observe estimates $V_1(\omega), V_2(\omega), \ldots, V_j(\omega)$ of values

$$R\left(X_n(\omega) + \frac{a_n}{2}\, Y_n(\omega)\right),\ R\left(X_n(\omega) + 2\frac{a_n}{2}\, Y_n(\omega)\right), \ldots, R\left(X_n(\omega) + j\,\frac{a_n}{2}\, Y_n(\omega)\right)$$

until all but the last have the same sign as the estimate $-Y_n(\omega)$ of $R(X_n(\omega))$. According to Theorem (4.4) we then put $\alpha_n(\omega) = j\,\frac{a_n}{2}$. If the errors $\tilde{V}_i = V_i - R\left(X_n - i\,\frac{a_n}{2}\, Y_n\right)$ are independently and identically distributed with

[5]) By a_+ or a^+ and a_- or a^- we denote the positive and negative part of a respectively; $= a_+ + a_-$.

$\mathbf{E}$ sign $\tilde{V}_i = 0$, if they are also independent of $\mathscr{X}_n$, Y_n, then all the conditions of Theorem (4.4) are satisfied $\left(\text{with } m_i = R\left(X_n - i\,\frac{a_n}{2}\,Y_n\right)\right)$ and Assumption (4.5) is again satisfied with $C = 6$.

Now we shall study the behaviour of X_n under the assumptions accepted. Without loss of generality we may assume that $\Theta = 0$. Then we have $f(x) = x^2$, $D(x) = 2x$, $-\mathbf{M}_n(\mathscr{X}_n)\,D(X_n) = 2R(X_n)\,X_n$ is non-negative and thus (5.2.1) is satisfied with $B_n = \sqrt{2X_nR(X_n)}$, $b_n = 0$. The assumption (1.2.4) implies (5.2.3) with $d_n = \sigma^2$, $e_n = 0$; a fortiori (5.2.3) holds with $e_n = A + B$, $d_n = B(A + B) + \sigma^2$; we shall show that with these e_n, b_n (5.2.2) also holds. Indeed by (1.2.4) and (1.2.2) we have

$$\|\mathbf{M}_n(\mathscr{X}_n)\|^2 = R^2(X_n) \leqq |R(X_n)|\,(A\,|X_n| + B) \leqq AB_n^2 + B\,|R(X_n)|\,.$$

Now for $|x| \leqq 1$ we have $|R(x)| \leqq A + B$, for $|x| > 1$ we have $|R(x)| \leqq |x|\,|R(x)| \leqq B_n^2$; hence

$$\|\mathbf{M}(\mathscr{X}_n\|^2 \leqq AB_n^2 + B(A + B + B_n^2) = (A + B)\,B_n^2 + B(A + B) \leqq e_nB_n + d_n\,.$$

Since (5.2.4) follows from (1.2.3), all assumptions of Theorem (5.2) hold. Hence $f(X_n)$ converges to a random variable and there exists a sequence n_i such that $B_{n_i}^2 = X_{n_i}R(X_{n_i}) \to 0$. Hence we get

$$\begin{aligned}\lim X_n^2 &\to \{a^2;\, x_i \in E_1,\, x_i \to a,\, x_iR(x_i) \to 0\} = \\ &= \{0\} \cup \{a^2;\, x_i \in E_1,\, x_i \to a,\ R(x_i) \to 0\}\,.\end{aligned}$$

If moreover (1.2.1) holds, $R(x_i) \to 0$ implies, if $x_i \to a$, that $x_i \to 0$ and thus in this case, $X_n \to 0$.

(6.2) Suppose we again seek the point Θ at which a function R, defined on E_1, acquires its zero value, we have the sequence a_n satisfying (1.2.3), Y_n are again estimates of $-R(X_n)$, Assumption (4.5) holds, but we put

$$X_{n+1} = X_n + \alpha_n \operatorname{sign} Y_n\,. \tag{6.2.1}$$

Denoting sign $Y_n = \widehat{Y}_n$, $\widehat{R}(X_n) = \mathbf{E}_{\mathscr{X}_n}\widehat{Y}_n$, we deal with the usual Robbins-Monro approximation scheme for the function $\widehat{R}$. Automatically it satisfies conditions (1.2.4) and (1.2.2) and the meaning of condition (1.2.1) or (6.1.1) for $\widehat{R}$ is clear from the relation

$$\widehat{R}(X_n) = P_{\mathscr{X}_n}(Y_n > 0) - P_{\mathscr{X}_n}(Y_n < 0)\,.$$

We note that the procedure (6.2.1) for $\alpha_n = a_n$ was already studied by Blum [1].

(6.3) Note. If α_n are determined in the way described in Theorem (4.4), the procedure is related to that of Harry Kesten [13] in the following way: if Y_n and V_i take on only the values $-1, 1$, the two methods are identical. Generally, instead of estimating R at the points $X_n(\omega) + aY_n(\omega)$, $X_n +$

$+ 2aY_n(\omega), \ldots, X_n + \alpha_n(\omega) Y_n(\omega)$, Kesten's method takes observations at the points $X_n(\omega) + aY_n(\omega)$, $X_n + a(Y_n(\omega) - V_1(\omega)), \ldots, X_n(\omega) + a(Y_n(\omega) - V_1(\omega) - \ldots - V_j(\omega))$ (where $\alpha_n(\omega) = ja$; however there are differences between Kesten's and our notations).

7. The Kiefer-Wolfowitz method and its modifications. (7.1) Suppose that R, Y_n, a_n, c_n satisfy the conditions of Theorem (1.3), but require, instead of (1.3.1), the following weaker condition

$$\text{(7.1.1)} \qquad \underline{D}(x) \geqq 0 \quad \text{for} \quad x \leqq \Theta, \quad \bar{D}(x) \leqq 0 \quad \text{for} \quad x \geqq \Theta\,.$$

Choose a c, $0 < c < 1$ and define $f(x) = [(|x - \Theta| - c)^+]^2$.

Suppose further that α_n are chosen in such a way that Assumption (4.5) holds. Theorem (4.1) says that it does so if $\alpha_n = a_n$ as in Theorem (1.3). However Theorems (4.2) and (4.3) show other possibilities of the choice. Having observed $Y_n(\omega)$ we may take estimates $V_i(\omega)$ of R at the points $X_n(\omega) + i \frac{a_n}{d} Y_n(\omega)$ unless $V_1(\omega) > V_2(\omega) > \ldots > V_j(\omega) \leqq V_{j+1}(\omega)$ and put $\alpha_n(\omega) = j \frac{a_n}{d}$.[6]) If further $d = \sum_{k=1}^{\infty} k^2 \frac{1}{(k+1)!}$, if the errors $V_i - R\left(X_n + i \frac{a_n}{d} Y_n\right)$ are continuous identically and independently distributed and independent of $\mathscr{X}_n, Y_n$, then the conditions of Theorem (4.3) hold $\left(\text{with } m_i = R\left(X_n + i \frac{a_n}{d} Y_n\right)\right)$, which implies that Assumption (4.5) is satisfied with $C = \sum_{k=1}^{\infty} k^2 \frac{1}{k!}$.

We shall study the behaviour of X_n. Without loss of generality we may assume that $\Theta = 0$ and, since only limiting properties are of interest and $c_n \to 0$, that $c_n < c$ for every n. This assumption together with (1.3.1) implies that

$$\text{(7.1.2)} \qquad \mathbf{M}_n(\mathscr{X}_n) = \frac{R(X_n + c_n) - R(X_n - c_n)}{2c_n} \begin{cases} \geqq 0 & \text{if} \quad X_n \leqq -c\,, \\ \leqq 0 & \text{if} \quad X_n \geqq c \end{cases}$$

and hence that — since $D(x) = 2(|x| - c)^+ . \operatorname{sign} x$

$$\text{(7.1 3)} \qquad D(X_n)\, \mathbf{M}_n(\mathscr{X}_n) \leqq 0$$

and we may put

$$\text{(7.1.4)} \qquad B_n = \sqrt{- D(X_n)\, \mathbf{M}_n(\mathscr{X}_n)}\,.$$

---

[6]) However in practice we choose not a_n but $a_n' = \frac{a_n}{d}$ and we need not know the values of $a_n = da_n'$ but only the limiting properties of a_n which are that of the sequence a_n'.

Now we shall show that the assumptions of Theorem (5.2) hold. First, (5.2.1) holds with $b_n = 0$ as follows from (7.1.4). (1.2.4) implies (5.2.3) with $d_n = \frac{\sigma^2}{2c_n}$, $e_n = 0$. Concerning (5.2.2) we get by (1.3.2) and since $c_n < c < 1$

$$\mathbf{M}_n^2(\mathscr{X}_n) \leqq \frac{|R(X_n + c_n) - R(X_n - c_n)|^2}{4c_n^2} \leqq$$
$$\leqq \frac{|R(X_n + c_n) - R(X_n - c_n)|(A|X_n| + B)}{2c_n^2} \leqq$$
$$\leqq \frac{|R(X_n + c_n) - R(X_n - c_n)|\,(A(|X_n| - c_n)^+ + A + B)}{2c_n^2} \leqq$$
$$\leqq \frac{A}{c_n} B_n^2 + \frac{A + B}{2c_n^2} |R(X_n + c_n) - R(X_n - c_n)| .$$

However $|R(X_n + c_n) - R(X_n - c_n)|$ is less than or equal to $4A + 2B$ or $(|X_n| - c_n)^+ |R(X_n + c_n) - R(X_n - c_n)|$ if $(|X_n| - c_n)^+ \leqq 1$ or $\geqq 1$ respectively. Hence

$$\mathbf{M}_n^2(\mathscr{X}_n) \leqq \frac{A}{c_n} B_n^2 + \frac{A + B}{c_n} B_n^2 + \frac{(A + B)(4A + 2B)}{2c_n^2}$$

and (5.2.3) is satisfied with $e_n = \frac{2A + B}{c_n}$ and $d_n = \frac{(A + B)(2A + B)}{c_n^2}$; both (5.2.2) and (5.2.3) are satisfied with $e_n = \frac{2A + B}{c_n}$, $d_n = \frac{(A + B)(2A + B) + \sigma^2}{c_n^2}$.

Concerning (5.2.4), the requirement $\sum a_n = +\infty$ is contained in (1.3.3), $\sum a_n b_n = 0$ since $b_n = 0$, $\sum a_n^2 d_n < +\infty$ since $\sum \frac{a_n^2}{c_n^2} < +\infty$ by (1.3.3). From the last inequality it follows that $\left(\frac{a_n}{c_n}\right)^2 \to 0$; hence $\frac{a_n}{c_n} \to 0$, too, $\lim a_n e_n = 0$ and (5.2.4) holds.

Since $B_n^2 = |D(X_n)|\,|\mathbf{N}_n(X_n)|$, where $\mathbf{N}_n(X_n) = \frac{R(X_n + c_n) - R(X_n - c_n)}{2c_n}$ and $\inf\limits_{|x| > c+\varepsilon} |D(x)| > 0$ for every $\varepsilon > 0$, we deduce from Theorem (5.2) that there exists a set $\Omega(c)$ such that $\Omega(c) \subset \Omega$, $P(\Omega(c)) = 1$ and that for every $\omega \in \Omega(c)$

(7.1.5) $\lim (|X_n(\omega)| - c)_+^2$ exists and (equals zero or $\mathbf{N}_{n_i}(X_{n_i}(\omega)) \to 0$) .

Since c was an arbitrary positive number (7.1.5) holds for every $\omega \in \Omega_0 = \bigcap_{k=1}^{\infty} \Omega\left(\frac{1}{k}\right)$ and every $c > 0$, which implies that

(7.1.6) $\lim X_n^2(\omega)$ exists and (equals zero or $\mathbf{N}_{n_i}(X_{n_i}(\omega)) \to 0$)

for every $\omega \in \Omega_0$, where $P(\Omega_0) = 1$. Obviously, if (1.3.1) holds, then $\lim X_n(\omega) = 0$ for every $\omega \in \Omega_0$.

(7.2) Suppose we again seek for the maximum of a non-negative function R, defined on E_1, but now we define Y_n to be $\frac{1}{2c_n}$ sign $(Y_n^+ - Y_n^-)$, where Y_n^+ and Y_n^- are estimates of $R(X_n + c_n)$ and $R(X_n - c_n)$ respectively. Suppose that as in Theorem (1.3) $\mathbf{E}_{\mathscr{X}_n} Y_n = \mathbf{M}_n(\mathscr{X}_n) = \mathbf{N}_n(X_n)$ is a function of X_n only and suppose (instead of (1.3.1) or (7.1.1)) that

(7.2.1) $$\mathbf{N}_n(x) \geqq 0 \text{ for } x < \Theta - c_n,\ \mathbf{N}_n(x) \leqq 0 \text{ for } x > \Theta + c_n\ .$$

The conditions (1.3.2) and (1.3.4) will be omitted. Further we suppose that (1.3.3) holds and that α_n satisfies Assumption (4.5) with $f(x) = [(|x - \Theta| - c)^+]^2$ for every $0 < c < 1$.

Under these conditions we shall study the behaviour of X_n. As in (7.1) we suppose that $\Theta = 0$, $c_n < c$ for every n. According to (7.2.1) $\mathbf{M}_n(\mathscr{X}_n)\, D(X_n)$ is non-positive, so that (5.2.1) holds with $b_n = 0$ and $B_n = \sqrt{- \mathbf{N}_n(X_n)\, D(X_n)}$. Since $|Y_n| \leqq \frac{1}{c_n}$, we have (5.2.2) and (5.2.3) with $d_n = \frac{1}{c_n}$, $e_n = 0$. (5.2.4) follows easily from (1.3.3) and from the relations $e_n = 0$, $b_n = 0$, $d_n = \frac{1}{c_n}$. Since f satisfies Assumption (2.1), we get from Theorem (5.2), the conditions of which we have already verified, that there exists a $\Omega(c) \subset \Omega$ such that $P(\Omega(c)) = 1$ and that for every ω in $\Omega(c)$ (7.1.5) holds. Putting $\Omega_0 = \bigcap_{k=1}^{\infty} \Omega\left(\frac{1}{k}\right)$, we get that (7.1.6) holds for every $\omega \in \Omega_0$ and that $P(\Omega_0) = 1$.

If instead of (7.2.1) the following stronger condition

(7.2.7) $$\begin{aligned} &\inf\,\{\mathbf{N}_n(x);\ n = 1, 2, \ldots, x \in (-n, -c_n + \Theta)\} > 0\ , \\ &\sup\,\{\mathbf{N}_n(x);\ n = 1, 2, \ldots, x \in (\Theta + c_n, n)\} < 0 \end{aligned}$$

is satisfied, then obviously $X_n \to \Theta$.

8. Multidimensional case. (8.1) Suppose that R, Y_n, a_n, c_n satisfy the conditions of Theorem (1.4) with the exception of (1.4.1) and that α_n satisfies Assumption (4.5) with $f = R$. (By Theorem (4.1) the last condition is satisfied if the α_n are chosen as in Theorem (1.4); it is also satisfied if the α_n are determined in the way described in Theorem (4.2) resp. (4.3) — see also (7.1)). Under these conditions we shall study the behaviour of X_n.

From (1.4.4) it follows by Taylor's Theorem that

$$\mathbf{M}_n(\mathscr{X}_n) = - DR(X_n) - \frac{c_n}{2} D_2 R\left(\chi(X_n)\right),$$

where $\chi^{(i)}(X_n) \in (X_n^{(i)}, X_n^{(i)} + c_n)$. According to (1.4.2) the assumptions of Theorem (5.4) hold with $H_n(\mathscr{X}_n) = 1$, $h_n = Kc_n$, $\Theta_n(\mathscr{X}_n) = - \frac{D_2R(\chi(X_n))}{2K}$,

$g_n = 1, C_1 = 0$. Thus (5.2.1) holds for $B_n = \|D(X_n)\|$, $b_n = Kc_n$, $K_2 = 0$. Further $\|\mathbf{M}_n(\mathscr{X}_n)\| = \|D(X_n)\|^2 + 2h_n \langle D(X_n), \Theta_n(\mathscr{X}_n)\rangle + h_n^2 \|\Theta_n(\mathscr{X}_n)\| \leqq B_n^2 + 2h_nB_n + h_n^2 \leqq (1 + 2h_n) B_n^2 + 2h_n + h_n^2$, whence it follows that (5.2.2) holds with $e_n = 1 + 2Kc_n$, $d_n = 2Kc_n + 2K^2c_n^2 + \frac{\sigma^2}{c_n^2}$; from (1.4.5) it follows that (5.2.3) holds with these c_n, d_n, too.

Concerning (5.2.4): the condition $\sum a_n = +\infty$ is contained in (1.4.3); $\sum a_nb_n < +\infty$ is satisfied since by (1.4.3) $\sum a_nc_n < +\infty$ and $b_n = Kc_n$; $\sum a_nd_n < +\infty$ follows from the relations (see (1.4.3)) $c_n \to 0$, $\sum \frac{a_n^2}{c_n^2} < +\infty$, which imply $d_n = 2Kc_n + 2K^2c_n^2 + \frac{\sigma^2}{c_n^2} < \frac{2\sigma^2}{c_n^2}$ for large n. The relations $\lim b_n = \lim a_ne_n = 0$ follow from the assumptions $\sum a_nc_n < +\infty$, $\sum \frac{a_n^2}{c_n^2} < +\infty$, $c_n \to 0$ which imply $a_n \to 0$ and from the relations $b_n = Kc_n$, $e_n = 1 + 2Kc_n$. Obviously f satisfies Assumption (2.1) and by Theorem (5.2) there exists a set $\Omega_0 \subset \Omega$ with probability one such that for every $\omega \in \Omega_0$ $\lim R(X_n(\omega))$ exists and belongs to the set

$$R\,\{x; DR(x) = 0\} \cup \{a; x_i \in \mathbf{X}, \|x_i\| \to +\infty, R(x_i) \to a, DR(x_i) \to 0\}\,.$$

For the interpretation of this result see Note (5.3).

Now we shall study the modification of the choice of Y_n, analogous to those investigated in sections (6.2) and (7.2). There, under some conditions on the observations of function considered, the modification enabled us to omit conditions (1.2.2) and (1.3.2), respectively. Here we shall give some conditions on Y_n sufficient to ensure that the convergence will not break down (Theorem (8.4)) and that even under some other conditions the condition (1.4.2) can be weakened (Theorem (8.5)). First we shall state an assumption.

(8.2) Assumption. G is a distribution function with a bounded continuous derivative g, σ is a positive bounded function on $\mathbf{X}$, $\frac{1}{\sigma}$ is bounded and has a continuous derivative $D\frac{1}{\sigma}$, R is a function on $\mathbf{X}$ with a continuous derivative DR. For every principal n, c_n is a positive number, $Z_{n,+}$, $Z_{n,-}$ are random vectors,

$$Y_n^{(i)} = -\frac{1}{c_n}\operatorname{sign}(Z_{n,+}^{(i)} - Z_{n,-}^{(i)}) \quad (i = 1, \ldots, q)\,, \tag{8.2.1}$$

$$P_{\mathscr{X}_n}\{Z_{n,-}^{(i)} \leqq z\} = G\left(\frac{z - R(X_n)}{\sigma(X_n)}\right) \quad (i = 1, \ldots, q)\,, \tag{8.2.2}$$

$$P_{\mathscr{X}_n}\{Z_{n,+}^{(i)} \leqq z\} = G\left(\frac{z - R(X_n + c_n\Delta^{(i)})}{\sigma(X_n + c_n\Delta^{(i)})}\right) \quad (i = 1, \ldots, q)\,, \tag{8.2.3}$$

$Z_{n,+}^{(i)}$, $Z_{n,-}^{(i)}$ are conditionally ($\mathscr{X}_n$) independent, i. e.

(8.2.5) $P_{\mathscr{X}_n}\{Z^{(i)}_{n,-} \leqq z_1,\ Z^{(i)}_{n,+} \leqq z_2\} = P_{\mathscr{X}_n}\{Z^{(i)}_{n,-} \leqq z_1\} \,.\, P_{\mathscr{X}_n}\{Z^{(i)}_{n,n} \leqq z_2\}$

for every $z_1, z_2 \in E_1$, $i = 1, \ldots, q$

and either

(8.2.6) $$\sigma = 1$$

or

(8.2.7) $$\int_{-\infty}^{+\infty} y g^2(y)\,\mathrm{d}y = 0\,.$$

(8.3) Lemma. *Let Assumption* (8.2) *holds. Then*

(8.3.1) $$\mathbf{M}^{(i)}_n(\mathscr{X}_n) = -\frac{1}{c_n}[1 - 2r_i\,(X_n, c_n)]$$

where

(8.3.2) $$r_i(x, 0) = \frac{1}{2}\,,$$

(8.3.3) $$r_i(x, c) = \frac{1}{\sigma(x)}\int_{-\infty}^{+\infty} G\left(\frac{y - R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})}\right) g\left(\frac{y - R(x)}{\sigma(x)}\right)\mathrm{d}y\,,$$

(8.3.4) $$\frac{\mathrm{d}}{\mathrm{d}c}\, r_i(x, c)\Big/_{c=0} = -\frac{D^{(i)}R(x)}{\sigma(x)}\int_{-\infty}^{+\infty} g^2(y)\,\mathrm{d}y$$

and, if $\sigma = 1$,

(8.3.5)
$$\frac{\mathrm{d}}{\mathrm{d}c}\, r_i(x, c) = -D^{(i)}R(x + c\Delta^{(i)})\int_{-\infty}^{-\infty} g(y - R(x + c\Delta^{(i)})\, g(y - R(x)))\,\mathrm{d}y\,.$$

Proof. As follows from the definition of Y_n, (8.3.1) will be satisfied if

(8.3.6) $$r_i(X_n, c_n) = P_{\mathscr{X}_n}(Z^{(i)}_{n,+} - Z^{(i)}_{n,-} \leqq 0)\,.$$

From (8.2.2) to (8.2.5) it follows that

$$P_{\mathscr{X}_n}(Z_{n,+} - Z_{n,-} \leqq z) = \int_{-\infty}^{+\infty} G\left(\frac{z - R(X_n + c_n\Delta^{(i)}) - y}{\sigma(X_n + c_n\Delta^{(i)})}\right).\,\mathrm{d}G\left(\frac{-R(X_n) - y}{\sigma(X_n)}\right),$$

whence, substituting $z = 0,\ -y = t$,

(8.3.7) $$r_i(x, c) = \int_{-\infty}^{+\infty} G\left(\frac{t - R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})}\right)\mathrm{d}G\left(\frac{t - R(x)}{\sigma(x)}\right)$$

which is equivalent to (8.3.3).

The relation (8.3.2) follows from the fact that $r(x, 0)$ equals $P\{V_1 - V_2 \leqq \leqq 0\}$ for two independent continuous and identically distributed (with distribution function $G\left(\frac{(v - R(x)}{\sigma(x)}\right)$, random variables V_1, V_2. Differentiating the integrand in (8.3.3) gives

$$- g\left(\frac{y - R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})}\right) g\left(\frac{y - R(x)}{\sigma(x)}\right)\left[\frac{D^{(i)}R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})} + \right.$$
$$\left. + [y - R(x + c\Delta^{(i)})]\, D^{(i)} \frac{1}{\sigma(x + c\Delta^{(i)})}\right];$$

from Assumption (8.2) we deduce easily, that this expression has for every given $x \in \mathbf{X}$ and c in every finite interval (c_1, c_2) a integrable majorante. Thus we may differentiate under the sign of the integral in (8.3.3). If $\sigma = 1$, we have $D^{(i)} \frac{1}{\sigma} = 0$ and (8.3.5) holds. If $c = 0$, then

$$\frac{\mathrm{d}r_i(x, c)}{\mathrm{d}c}\bigg/_{c=0} = - \frac{D^{(i)}R(x)}{\sigma^2(x)} \int_{-\infty}^{+\infty} g^2\left(\frac{y - R(x)}{\sigma(x)}\right) \mathrm{d}y -$$
$$- D^{(i)} \frac{1}{\sigma(x)} \int_{-\infty}^{+\infty} \frac{y - R(x)}{\sigma(x)}\, g^2\left(\frac{y - R(x)}{\sigma(x)}\right) \mathrm{d}y =$$
$$= - \frac{D^{(i)}R(x)}{\sigma(x)} \int_{-\infty}^{+\infty} g^2(y)\, \mathrm{d}y - \sigma(x)\, D^{(i)} \frac{1}{\sigma(x)} \int_{-\infty}^{+\infty} y g^2(y)\, \mathrm{d}y\,.$$

Hence the relation (8.3.4) follows either by (8.2.6) or (8.2.7) and the proof is accomplished.

(8.4) **Theorem.** *Let $f = R$ satisfy Assumptions* (2.1) *and* (4.5), *let the random variables Y_n satisfy Assumption* (8.2) *with $\sigma = 1$, let the positive numbers a_n, c_n satisfy the relations*

(8.4.1) $\quad \sum a_n = +\infty, \quad \sum a_n c_n^2 < +\infty, \quad \sum \frac{a_n^2}{c_n^2} < +\infty, \quad \lim c_n = 0$[7])

Then for almost all ω $\lim R(X_n(\omega))$ exists and belongs to the set $A = A_1 \cup A_2$, where

$$A_1 = R\,\{x;\, DRx = 0\}\,,$$
$$A_2 = \{a;\, a \in E_1,\, x_i \in \mathbf{X},\, R(x_i) \to a,\, \|x_i\| \to +\infty\}\,.$$

Proof. By the Mean Value Theorem we get

$$\mathbf{M}_n^{(i)}(\mathscr{X}_n) = - \frac{1}{c_n} + \frac{2}{c_n}\, r_i\,(X_n, c_n) =$$
$$= - \frac{1}{c_n} + \frac{2}{c_n}\left[r_i(X_n, 0) + c_n \frac{\mathrm{d}}{\mathrm{d}c}\, r_i(X_n,\, \Theta_i(X_n))\right]$$

[7]) This is a rather weaker condition than (1.4.3).

with $0 < \Theta_i(X_n) < c_n$, since by Lemma (8.3) the derivative of $r_i(x, c)$ exists. However by (8.3.2) $r_i(X_n, 0) = \frac{1}{2}$ and thus according to (8.3.5)

$$(8.4.2) \qquad \mathbf{M}_n^{(i)}(\mathscr{X}_n) = - D^{(i)}R(X_n + \Theta_i(X_n)\,\Delta^{(i)})\,\varkappa_i(X_n)\,,$$

where

$$(8.4.3) \quad \varkappa_i(X_n) = \int_{-\infty}^{+\infty} g[y - R(X_n + \Theta_i(X_n)\,\Delta^{(i)})]\,.\,g(y - R(X_n))\,\mathrm{d}y \geqq 0\,.$$

According to (8.4.2) $\mathbf{M}_n^{(i)}(\mathscr{X}_n)\,D^{(i)}R(X_n)$ is non-positive if $D^{(i)}R(x) \neq 0$ for every $x \in (X_n, X_n + c_n\Delta^{(i)})$. In the opposite case, since $\|D_2R\| < 2K$, $|D^{(i)}R(x)| < 2Kc_n$ for every $x \in (X_n, X_n + c_n\Delta^{(i)})$. Thus

$$\mathbf{M}_n^{(i)}(\mathscr{X}_n)\,D^{(i)}R(X_n) \leqq 4K^2c_n^2\varkappa_i(X_n)\,.$$

By Assumtion (8.3) g is bounded. Hence $\varkappa(X_n)$ is also bounded and we get

$$(8.4.4) \qquad \langle\mathbf{M}_n(\mathscr{X}_n),\ D(X_n)\rangle \leqq - B_n^2 + c_n^2K_2\,,$$

with a suitable constant K_2 and with

$$(8.4.5) \qquad B_n^2 = - (\langle\mathbf{M}_n(\mathscr{X}_n),\ D(X_n)\rangle)_-\,.$$

Now we shall apply Theorem (5.2). (8.4.4) shows that (5.2.1) is satisfied with $b_n = c_n^2$. From the definition of Y_n it follows that both (5.2.2) and (5.2.3) are satisfied with $e_n = 0$, $d_n = \frac{q^2}{c_n^2}$. Thus the condition (5.2.4) can be rewritten as $\sum a_n = +\infty$, $\sum a_n c_n^2 < +\infty$, $\sum \frac{a_n^2}{c_n^2} < +\infty$, $\lim c_n^2 = 0$ and these relations are assumed in (8.4.1). Thus all conditions of Theorem (5.2) are satisfied and thus for almost all ω in Ω $\lim R(X_n(\omega))$ exists and $B_{n_j}(\omega) \to 0$, i. e. $\langle\mathbf{M}_{n_j}(\mathscr{X}_{n_j}(\omega)), D(X_{n_j}(\omega))\rangle_- \to 0$ for a sequence of natural numbers n_j.

However the positive part of $\langle\mathbf{M}_{n_j}(\mathscr{X}_{n_j}(\omega)),\ D(X_{n_j}(\omega))\rangle$ converges to zero by (8.4.4), too and

$$\langle\mathbf{M}_{n_j}(\mathscr{X}_{n_j}(\omega)),\ D(X_{n_j}(\omega))\rangle \to 0. \quad \text{Thus for every } i = 1, 2, \ldots, q$$

$$D^{(i)}R(X_{n_j}(\omega))\ D^{(i)}R(X_{n_j}(\omega) + \Theta_i(X_n(\omega))\Delta_i)\ \varkappa_i(X_n(\omega)) \to 0\,.$$

Thus there exist $x_j, x_j' \in \mathbf{X}$, such that $|x_j - x_j|' \to 0$ and that $R(x_j) \to a = \lim R(X_n(\omega))$,

$$[D^{(i)}R(x_j)\ D^{(i)}R(x_j')] \int_{-\infty}^{+\infty} g(y - R(x_j'))\,g(y - R(x_j))\,\mathrm{d}y \to 0\,.$$

Since from Assumption (2.1) it follows that $D^{(i)}R$ is uniformly continuous, the last relation is satisfied only if

$$D^{(i)}R(x_j) \to 0 \quad \text{or} \quad \int_{-\infty}^{+\infty} g(y - R(x_j'))\,g(y - R(x_j))\,\mathrm{d}y \to 0\,.$$

Now if the sequence $\|x_j\|$ is not bounded, it is easy to see (by taking such a subsequence x_{n_j} that $\|x_{n_j}\| \to +\infty$) that $a \in A_2$. If $\|x_j\| < M$ for some M, then from the continuity of DR there follows the uniform continuity of R in the sphere $\{x; \|x\| < 2M\}$ and $R(x'_j) - R(x_j) \to 0$. By boundedness and continuity of g and R we get

$$\int_{-\infty}^{+\infty} g(y - R(x'_j)) g(y - R(x_j)) \, dy =$$

$$= \int_{-\infty}^{+\infty} g(y - (R(x'_j) - R(x_j)) g(y) \, dy \to \int_{-\infty}^{+\infty} g^2(y) \, dy > 0 \, .$$

Thus if a non $\in A_2$, then $D^{(i)}R(x_j) \to 0$ and there exists a subsequence x_{n_j} converging to a point $x \in X$ such that we get $\lim R(X_n(\omega)) = \lim R(x_{n_j}) = R(x) \in A_1$ since $DR(x) = \lim DR(x_{n_j}) = 0$; the proof is finished.

Remark. If $|R(x_j)| \to +\infty$ as soon as $\|x_i\| \to +\infty$ then $A_2 = \emptyset$. If Theorem (5.5) can be applied, we get $\sup X_n(\omega) < +\infty$ with probability one and the sequence x_j in the proof of the preceding sequence can be supposed to be bounded, whence again we get $\lim R(X_n(\omega)) \in A_1$ with probability one.

(8.5) Theorem. *Suppose that R is a function ou $\mathbf{X}$ with a second derivative. Let ϱ be a function defined on E_1 with a derivative ϱ' satisfying*

(8.5.1) $$\inf_{x \in A} \varrho'(x) > 0 \text{ for every bounded set } A \subset E_1 \, ,$$

suppose that Assumptions (8.2), (2.1) *and* (4.5) *are satisfied with $f = \varrho(R)$ and with*

(8.5.2) $$\sum a_n = +\infty, \ \sum a_n c_n < +\infty, \ \sum \frac{a_n^2}{c_n^2} < +\infty, \ c_n \to 0 \, .$$

Suppose that for every $x \in \mathbf{X}$ there exists a function φ_x defined on E_1 and a positive number $c(x)$ such that for every $c \in (0, c(x))$, $y \in E_1$ we have

8.5.3) $$\left| \frac{d^2}{dc^2} G \left(\frac{y - R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})} \right) \right| \leqq \varphi_x(y)$$

and for every $x \in \mathbf{X}$

(8.5.4) $$\int_{-\infty}^{+\infty} \varphi_x(y) \, g(y) \, dy < +\infty \, .$$

Finally suppose that there exist such positive constants K_2, γ that

(8.5.5) $$\left| \varrho'(R(x)) \, D^{(i)}R(x) \int_{-\infty}^{+\infty} \frac{d^2}{dc^2} G \left(\frac{y - R(x + c\Delta^{(i)})}{\sigma(x + c\Delta^{(i)})} \right) \cdot g \left(\frac{y - R(x)}{\sigma(x)} \right) dy \right| < K_2$$

for every $x \in \mathbf{X}$, $c \in (0, \gamma)$.

Then for almost all ω $\lim_{n \to \infty} R(X_n(\omega))$ exists (possibly infinite) und belongs to the set $A = A_1 \cup A_2 \cup A_3$, where

(8.5.6) $$A_1 = R(\{x; DR(x) = 0\})$$

(8.5.7) $A_2 = \{a; x_i \in \mathbf{X}, \|x_i\| \to +\infty, R(x_i) \to a \in E_1, DR(x_i) \to 0\}$

(8.5.8) $A_3 = \{a; x_i \in \mathbf{X}, \|x_i\| \to +\infty, |R(x_i)| \to +\infty\}$.

If $P\{\sup X_n(\omega) < +\infty\} = 1$[8]) *then* $P\{\lim R(X_n(\omega)) \in A_1\} = 1$.

Remark. The meaning of conditions (8.5.3) and (8.5.4) is clear: they ensure the possibility of differentiating twice under the sign of integral in (8.3.3). It can be easy seen that they will be satisfied if e. g. R and $\frac{1}{\sigma}$ have continuous second derivatives and if G has a bounded second derivative.

If we use Theorem (5.2), then the function f, which can be said to measure the success of approximation, must satisfy Assumtion (2.1). One way of choosing f is to put $f = R$, as we have done in section (8.1); then we must require that R is upper bounded and has a bounded second derivative. These last conditions can be weakened by the introduction of an increasing function ϱ.

If we put for example

$$\varrho(y) = \begin{cases} e^y & \text{for } y \leqq 0, \\ 2e - e^{-y} & \text{for } y > 0 \end{cases}$$

then Assumption (2.1) is satisfied for $f = \varrho(R)$ if R is a polynomial of any degreee.

Condition (8.5.5) will be satisfied, too, for a large class of functions R, for which $\|DR(x_i)\| \to +\infty$ or $\|D_2R(x_i)\| \to \infty$ implies $|R(x_i)| \to \infty$ and for a suitable ϱ. Indeed, if for simplicity we assume $\sigma = 1$, the condition (8.5.5) can be written as

$$x \in \mathbf{X},\ c \in (0, \gamma),\ i = 1, 2, \ldots, q \Rightarrow \Big| \int_{-\infty}^{+\infty} \{g'(y - R(x + c\Delta^{(i)}))\,[DR(x + c\Delta^{(i)})]^2 +$$
$$+ g(y - R(x + c\Delta^{(i)})\, D_2^{(ii)}R(x + c\Delta^{(i)})\}\, g(y - R(x)\, \mathrm{d}y\varrho'(R(x))\, D^{(i)}R(x)\Big| < K_2$$

i. e.

$$|\varrho'(R(x)\,[D^{(i)}R(x + c\Delta^{(i)})]^2\, D^{(i)}R(x) \int g'(y - R(x + c\Delta^{(i)})]\,.\,g(y - R(x))\,\mathrm{d}y +$$
$$+ \varrho'(R(x))\, D_2^{(ii)}R(x + c\Delta^{(i)})\, D^{(i)}R(x)\,.\int g(y - R(x + c\Delta^{(i)}))\, g(y - R(x)\,\mathrm{d}y| < K_2\,.$$

It is easy to see that the last inequality will be satisfied again if g' is bounded, ϱ defined as above and if R is a polynomial, K_2 and γ a suitable positive number.

Proof. Since Assumption (8.2) is satisfied, we may use Lemma (8.3). (8.5.3) ensures that we may integrate twice under the sign of integration in (8.3.3) and according to (8.3.1), (8.3.2) and (8.3.4) we get by Taylor's Theorem

(8.5.9) $\mathbf{M}_n(\mathscr{X}_n) = -h(X_n)\, DR(X_n) - c_n\Theta_n(X_n)$,

---

[8]) See Theorem (5.5).

where

$$h(x) = \frac{2}{\sigma(x)} \int_{-\infty}^{+\infty} g^2(y) \, \mathrm{d}y \tag{8.5.10}$$

and

$$\Theta_n^{(i)}(X_n) = \frac{1}{\sigma(x)} \int_{-\infty}^{+\infty} \frac{\mathrm{d}^2}{\mathrm{d}c^2} G\left(\frac{y - R(X_n + c\Delta^{(i)})}{\sigma(X_n + c\Delta^{(i)})}\right) \cdot g\left(\frac{y - R(X_n)}{\sigma(X_n)}\right) \mathrm{d}y \tag{8.5.11}$$

with $0 < c < c_n$.

Now since $f = \varrho(R)$ we have

$$D(x) = \varrho'(R(x)) \, DR(x) \tag{8.5.12}$$

and thus according to (8.5.9) and (8.5.5) for sufficiently large n

$$\langle \mathbf{M}_n(\mathscr{X}_n), D(X_n)\rangle = -\varrho'(R(X_n)) \, h(X_n) \, \|DR(X_n)\|^2 - c_n\varrho'(R(X_n)) < D(R(X_n)) \, , \, \Theta_n(X_n) > \leqq -\varrho'(R(X_n)) \, h(X_n) \, \|DR(X_n)\|^2 + c_n K$$

for a suitable constant K so that (5.2.1) is satisfied with

$$B_n^2 = B^2(X_n) = \varrho'(R(X_n)) \, h(X_n) \, \|DR(X_n)\|^2 \, , \, b_n = c_n \, . \tag{8.5.13}$$

Clearly both (5.2.2) and (5.2.3) are satisfied with $e_n = 0$, $d_n = \frac{q^2}{c_n^2}$, so that (5.2.4) follows from (8.5.2) and all the assumption of Theorem (5.2) are satisfied. Hence for almost all $\omega \in \Omega$ $\lim \varrho(R(X_n(\omega)))$ exists and belongs to the set defined in (5.2.5). Since ϱ is increasing, $\lim R(X_n(\omega))$ also exists, however is not necessarily finite. If the sequence $X_n(\omega)$ is bounded then there exists a subsequence n_i such that for $x_i = X_{n_i}(\omega)$ we have $x_i \to x \in \mathbf{X}$, $R(x_i) \to R(x)$, $B_{n_i}^2(\omega) = \varrho'(R(x_i)) \, h(x_i) \, \|DR(x_i)\|^2 \to 0$. However the sequence $R(x_i)$ is bounded, σ is bounded and by (8.5.1) and (8.5.10) we get $DR(x_i) \to 0$. Since $DR(x_i) \to DR(x)$, $\lim R(X_n(\omega)) \in A_1$.

If $X_n(\omega)$ is not bounded but $R(X_n(\omega))$ is so, then again from $B_{n_i}(\omega) \to 0$ it follows that $DR(X_{n_i}(\omega)) \to 0$ and $\lim R(X_n(\omega)) \in A_2$. If neither $X_n(\omega)$ nor $R(X_n(\omega))$ are bounded then $\lim R(X_n(\omega)) \in A_3$.

9. Concluding Remarks. (9.1) Other definitions of Y_n. To observe Y_n considered in the two last sections it suffices to take estimates of $R(x)$ at the points

$$X_n(\omega), \quad X_n(\omega) + c_n\Delta^{(i)}, \quad i = 1, \ldots, q \, ,$$

i. e. to take $q + 1$ observations of random variables. Jerome Sacks [16] points out that this definition of Y_n leads to a systematical bias of X_{n+1} — considered as an estimate of Θ in Theorem (1.4) — and propose to estimate $R(x)$ at the $2q$ points $X_n \pm c_n\Delta^{(i)}$. However since this bias is known, an estimate of Θ can be obtained without increasing the number of observations.

On the other hand there may be many other possibilities of the choice of Y_n. For example if $q = 3$ we may use a Latin square $2 . 2 = q + 1$, observe the estimates V_{ij} of $R(X_{ij})$, where

$$X_{ij}^{(1)} = \begin{cases} X_n^{(1)} - c_n \Delta^{(1)} & \text{for } i = 1 , \\ X_n^{(1)} + c_n \Delta^{(1)} & \text{for } i = 2 , \end{cases} \qquad X_{ij}^{(2)} = \begin{cases} X_n^{(2)} - c_n \Delta^{(2)} & \text{for } j = 1 , \\ X_n^{(2)} + c_n \Delta^{(2)} & \text{for } j = 2 , \end{cases}$$

$$X_{ij}^{(3)} = \begin{matrix} X_n^{(3)} - c_n \Delta^{(3)} & \text{for } i = j , \\ X_n^{(3)} + c_n \Delta^{(3)} & \text{for } i \neq j \end{matrix}$$

define Z_n by

$$Z_n^{(1)} = (V_{11} - V_{21}) + (V_{12} - V_{22}) , \qquad Z_n^{(2)} = (V_{11} - V_{12}) + (V_{21} - V_{22}) ,$$
$$Z_n^{(3)} = (V_{11} + V_{22}) - (V_{12} + V_{21})$$

and put $Y_n = \frac{1}{4c_n} Z_n$ (as an alogue to the definition consideied in Sec. (8.1)) or $Y_n^{(i)} = \frac{1}{4c_n} \operatorname{sign} Z_n^{(i)}$ (as analogue to the definition in Assumption (8.2)). It is easy to see that this definition of Y_n leads to no complications in proving the convergence properties of X_n under suitable conditions.

(9.2) Increasing the number of observations by increasing the dimension of **X**. The question often arising in practice if the process studies does or does not depend on a certain factor has the following abstract formulation. Given a function f on E_q does there exist a $\tilde{f}$ defined on E_{q-1} such that $f(x) = \tilde{f}(\tilde{x})$ for every $x \in E_q$, $\tilde{x} \in E_{q-1}$, $x^{(i)} = \tilde{x}^{(i)}$ for $i = 1, \ldots, q-1$? In the search for the minimum of f an erroneous positive answer to the preceding question results in reducing the number of observations but also in approximating the restricted $\inf\limits_{x^{(q)}=a} f(x)$, where a is a number, instead of approximating $\inf f(x)$. This error (of the first kind, say) can be of an essential character. The error of the second kind in answering our question in the negative leads to an increase in the number of observations. If the increase is large (and this is so for example if factorial designs are used with k levels for the q-th factor; then we need k times more observations), then the experimenter trying to avoid the Scylla of the perhaps unnecessary and large increase in the number of observations easily fails to avoid the Charybda and neglects practically significant factors. On the other hand a small increase diminishes this risk. And this is a further advantage of approximation methods desribed in Theorems (8.4) and (8.5), since there consideration of the function f defined on E_q instead of $\tilde{f}$ defined on E_{q-1} (if $f(x) = \tilde{f}(\tilde{x})$ as above) results in an increase in the number of observations at most by a factor $\frac{q+1}{q}$. Indeed if X_n and $\widehat{X}_n$

denote the approximation sequence for f and $\tilde{f}$ respectively, if the estimates are assumed to be equal in both processes as soon as the estimated quantities are identical, if further

$$X_n^{(i)} = \hat{X}_n^{(i)} \text{ for } i = 1, \ldots, q - 1 \tag{9.2.1}$$

and for $n = 1$, then it is easy to see that (9.2.1) holds for every $n = 1, 2, \ldots$. Hence our assertions follow from the fact that for the determination of the values of Y_n and $\hat{Y}_n$ we need $q + 1$ and q observations respectively and the number of observations for determining the value of α_n is identical in both cases.

(9.3.) Unsolved questions. From a host of them we mention especially two. The first was pointed already in Note (5.3): If in Theorem (5.5) $B = D$, under what non-trivial conditions the assertion $P\{\lim f(X_n) \in f\{x; D(x) = 0\}\}$ can be strengthened to $P\{\lim f(X_n) \in A\}$, where A is the set of local minima of f? Secondly how to generalize the consideration in a non-trivial way to functions f defined on a set $\mathbf{X} \subset E_q$ rather than on $\mathbf{X} = E_q$, especially if f may acquire its (possibly unique) minimum at the boundary of $\mathbf{X}$? Although we feel the great importance of the two problems we have not succeeded in solving them.

Acknowledgement. The author wishes to express his thanks to Mr. VL. JÁRA, the director and to Mr. J. SKŘIVÁNEK, the head of the research of the Research Institute for Inorganic Chemistry at Ústí nad Labem for enabling him to work at the problem while he was at the Institute.

BIBLIOGRAPHY

[1] *J. R. Blum:* Approximation methods which converge with probability one, AMS, *25* (1954), 382—386.

[2] *J. R. Blum:* Multidimensional stochastic approximation methods, AMS, *25* (1954), 463—483.

[3] *J. R. Blum:* A note on stochastic approximation, Proc. Am. Math. Soc. *9* (1958), 404—407.

[4] *G. E. P. Box, K. B. Wilson:* On the experimental attainment of optimum conditions, J. R. Stat. Soc., Series B, *13* (1951), 1—45.

[5] *C. Derman:* An application of Chung's lemma to the Kiefer Wolfowitz stochastic approximation procedure, AMS *27* (1956), 532—536.

[6] *J. L. Doob:* Stochastic processes, New York, 1953.

[7] *Václav Dupač:* O Kiefer-Wolfowitzové aproximační methodě, Čas. pěst. mat. *82* (1957), 47—75.

[8] *Václav Dupač:* Note on stochastic approximation methods, Čech. mat. žurn. *8* (*83*) (1958). 139—149.

[9] *Aryeh Dvoretzky:* On stochastic approximation, Proc. of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, Berkeley, 1956, 39—55.

[10] А. А. *Фельдбаум:* Автоматический оптимизатор, Автоматика и телемеханика, *19* (1958), 731—743.

[11] *K. L. Chung:* On a stochastic approximation method, AMS *25* (1954), 463—483.
[12] *L. B. Kantorovič:* On a effective method of solving extremal problems for quadratic functionals, Doklady A. N. SSSR *48* (1945), 455—460.
[13] *Harry Kesten:* Accelerated stochastic approximation, AMS *29* (1958), 41—59.
[14] *J. Kiefer, J. Wolfowitz:* Stochastic estimation of the maximum of a regresion function, AMS *23* (1952), 462—466.
[15] *H. Robbins, S. Monro:* A stochastic approximation method, AMS *22* (1951), 400—407.
[16] *Jerome Sacks:* Asymptotic distributions of stochastic approximations, AMS *29* (1958), 373—405.

Резюме

СТОХАСТИЧЕСКИЕ МЕТОДЫ ПРИБЛИЖЕНИЯ

ВАЦЛАВ ФАБИАН (Václav Fabian), Прага

Использование обычных схем $X_{n+1} = X_n + \alpha_n Y_n$ может оказаться практически невыгодным в случаях, когда $|E_{X_n} Y_n|$ велико для X_n близких и мало для X_n далеких от искомого решения. Этой невыгоды будут лишены схемы типа $X_{n+1} = X_n + \alpha_n \operatorname{sign} Y_n$.

Обычное предположение, что α_n — числа, может быть невыгодным в k-мерном случае при большом k, когда для определения направления Y_n необходимо произвести по меньшей мере $k + 1$ опытов. Так как неизвестна оптимальная длина шага в определенном таким образом направлении, представляется неэкономичным пробовать лишь одну длину, предписанную числом α_n. Определив направление Y_n, можно поступать, например, так (при разыскивании минимума функции R), что оцениваем последовательно $R(X_n + a_n Y_n)$, $R(X_n + 2a_n Y_n)$, ... при помощи оценок $V_1, V_2, \ldots$ до тех пор, пока не будет $V_1 > V_2 > \ldots > V_j \leqq V_{j+1}$, а затем можно положить $\alpha_n = j a_n$.

При довольно общих условиях, наложенных на оценки V_i, обычные аппроксимационные схемы сохраняют свою сходимость с вероятностью 1 при второй из указанных модификаций. Первая модификация также требует некоторого усиления условий, касающихся оценок функциональных значений, но зато позволяет ослабить условия, наложенные на регрессивные функции.

Свойства сходимости как модифицированных, так и исходных аппроксимационных схем, исследовались при более общих предположениях относительно регрессивных функций, чем, например, условия Й. Р. Блюма [2]. В случае отказа от условия (1.4.1) последовательность $R(X_n)$ сходится и ведет себя, грубо говоря, так, как будто бы X_n сходились к точке, в которой первая производная R равна нулю.