The Annals of Statistics
1988, Vol. 16, No. 3, 1330–1334

LOWER RATE OF CONVERGENCE FOR LOCATING A MAXIMUM OF A FUNCTION[1]

BY HUNG CHEN

State University of New York, Stony Brook

The problem is considered of estimating the point of global maximum of a function f belonging to a class F of functions on $[-1,1]$, based on estimates of function values at points selected possibly during the experimentation. If p is odd and greater than 1, K is a positive constant and F contains enough functions with pth derivatives bounded by K, then we prove that, under additional weak regularity conditions, the lower rate of convergence is $n^{-(p-1)/(2p)}$.

1. Introduction. The problem of designing an experiment to search for the location of a maximum of a function has been studied in stochastic approximation beginning with Kiefer and Wolfowitz (KW) (1952). A recent monograph is Nevelson and Hasminskii (1973). The stochastic approximation methods approximate a point of local maximum. Fabian (1967) described a modified KW procedure with convergence rate $\kappa_n = n^{-(p-1)/(2p)}$ for a class of functions with bounded pth derivatives, p an odd number.

The problem of locating a point of global maximum is treated in the response surface methodology of Box and Wilson (1951), in Nadaraya (1964) and Devroye (1978). Chen (1984) described a two-stage estimator, with rate of convergence $n^{-1/3}$, for estimating a point of global maximum within a class of functions with bounded third derivatives.

The question whether the convergence rates obtained are the best possible has been open (for almost 30 years). We shall show here that, indeed, the convergence rate κ_n cannot be improved for F containing enough functions with bounded pth derivatives.

Section 2 is preparatory. Condition 2.1 concerns the family F considered. Condition 2.3 is close to Cramér type conditions on densities in asymptotic considerations.

Definitions 2.4 and 2.5 specify the estimates considered.

Section 3 has the main result in Theorem 3.1. Remark 3.2 shows how the result can be extended to domains other than $[-1,1]$.

The result and the proof are close to those in Stone (1980, 1982). We treat a more specific case of the location of a maximum but allow for the design to depend on the observations; that is needed to include stochastic approximations.

2. Assumptions.

CONDITION 2.1. p is odd and greater than 1, $\kappa_n = n^{-(p-1)/(2p)}$ and δ_0 is a positive number. F is a family of functions on $[-1,1]$. For each f in F, $\chi(f)$ is

Received March 1985; revised January 1988.

[1]Work supported in part by NSF Grants MCS-83-01257 and MCS-80-02693.

AMS 1980 *subject classifications*. Primary 62G99; secondary 62L20.

Key words and phrases. Rate of convergence, global maximum.

a point of global maximum of f. For each $\delta \in [0, \delta_0]$ and every $n = 1, 2, \dots,$ the function

$$f_{n\delta}(x) = -x^2 + 2\delta n^{-1/2} \arctan\big(n^{1/(2p)}(x - \delta\kappa_n)\big) \tag{2.1}$$

is in F.

REMARK 2.2. Notice that

$$|f_{n\delta} - f_{n0}| \le \delta\pi n^{-1/2} \tag{2.2}$$

and that

$$\chi(f_{n0}) = 0, \qquad \chi(f_{n\delta}) = \delta\kappa_n \tag{2.3}$$

since

$$f'_{n\delta}(x) = -2x + 2\delta\kappa_n \arctan'\big(n^{1/(2p)}(x - \delta\kappa_n)\big).$$

For all δ sufficiently small, $f''_{n\delta} < -1$ and $|f^{(p)}_{n\delta}| \le 1$. In particular, such $f_{n\delta}$ are concave with a unique maximum.

CONDITION 2.3. μ is a measure on the σ-algebra of all Borel subsets of R_0. For each x and t (i.e., each x in $[0, 1]$ and t in R), $g(\cdot, x, t)$ is a probability density with respect to μ. For all y, x, t, $g(y, x, t) > 0$. For each y and x in R, the derivatives $g'(y, x, t)$ and $g''(y, x, t)$ with respect to t exist for all t. With $l = \log g$, we have, for all y, x, t,

$$|g''(y, x, t)| \le A(y, x), \qquad |l''(y, x, t)| \le B(y, x) \tag{2.4}$$

with functions A and B satisfying

$$\int A(\cdot, x)\, d\mu \le K, \qquad \int B(\cdot, x) g(\cdot, x, t)\, d\mu \le K \tag{2.5}$$

for a number K and all x, t.

DEFINITION 2.4. By a design we mean a rule that associates with each f in F two sequences $\langle X_n \rangle$ and $\langle Y_n \rangle$ of random variables, with the range of each X_n a subset of $[-1, 1]$ and such that the conditional distribution of X_n, given $X_1, \dots, X_{n-1}, Y_1, \dots, Y_{n-1}$ does not depend on f, and each Y_n has a conditional density with respect to μ, given $X_1, \dots, X_n, Y_1, \dots, Y_n$, equal to $g(\cdot, X_n, f(X_n))$.

We shall write Z_n for $\langle X_n, Y_n \rangle$ and $\mathscr{Z}_n$ for $\langle Z_1, \dots, Z_n \rangle$.

DEFINITION 2.5. By an estimate T we mean a pair $\langle \mathscr{D}, \langle T_n \rangle \rangle$ where $\mathscr{D}$ is a design and $T_1, T_2, \dots$ are random variables with each T_n a function of Z_n.

NOTATION 2.6. A design $\mathscr{D}$ and an f in F determine the distribution $P_{\mathscr{D}, f}$ of $\langle Z_n \rangle$. Since an estimate T specifies a design we will write also $P_{T, f}$ for $P_{\mathscr{D}, f}$.

3. The results.

THEOREM 3.1. *Under Conditions* 2.1 *and* 2.3, *for every number* $\eta \in (0, 1)$ *there exists a* $c > 0$ *such that, for all n and every estimate* T,

$$\inf_f P_{f, T}\{|T_n - \chi(f)| \ge c\kappa_n\} \ge \eta. \tag{3.1}$$

Proof. Let n be a positive integer and η a positive number smaller than 1. Set

$$C = \pi^2(K + \sqrt{K}), \qquad \varepsilon_0 = (1 - \eta)/10, \qquad \varepsilon = 2\varepsilon_0 \tag{3.2}$$

and choose δ_1 positive but small enough [cf. (2.4)] that $f_{n\delta}$ has a unique maximum at δ_{κ_n} for every n and every $\delta \le \delta_1$ (cf. Remark 2.2) and such that

$$\delta_1 \le \min\{\delta_0, \varepsilon_0^2/C\}. \tag{3.3}$$

Consider an estimate $\langle \mathscr{D}, \langle T_n \rangle \rangle$ and a δ in $(0, \delta_1]$ and set

$$f(x) = -x^2, \qquad f_n = f_{n,\delta}.$$

Denote $P_{\mathscr{D}, f}$ by P_0 and $P_{\mathscr{D}, f_n}$ by P_1 and denote by P_{0m}, P_{1m} the distribution of $\mathscr{Z}_m$ under P_0 and P_1, respectively. The density of P_{1n} with respect to P_{0n} is

$$L_n = \prod_{i=1}^{n} g(Y_i, X_i, f(X_i) + \tau_{ni})/g(Y_i, X_i, f(X_i)) \tag{3.4}$$

with

$$\tau_{ni} = f_n(X_i) - f(X_i). \tag{3.5}$$

Indeed, if (3.4) is true for n changed to $m - 1$, then L_{m-1} is also a density of Z_{m-1}, X_m under the two probability distributions, because the conditional distribution of X_m, given Z_{m-1}, is the same in the two cases. Then it is easy to see that (3.4) also holds for n replaced by m. Thus (3.4) holds.

Set $l_n = \log L_n$ and use Condition 2.3 to obtain

$$l_n = W_n + Z_n \tag{3.6}$$

with

$$W_n = \sum_{i=1}^{n} \tau_{ni} l'(Y_i, X_i, f(X_i)) \tag{3.7}$$

and

$$|Z_n| \le \tfrac{1}{2} \sum_{i=1}^{n} \tau_{ni}^2 B(Y_i, X_i). \tag{3.8}$$

By a standard argument [e.g., Fabian and Hannan (1985), Lemma 9.2.1 and the proof of Theorem 9.2.4], Condition 2.3 implies that the conditional expectation given Z_{i-1}, X_i, of $l'(Y_i, X_i, f(X_i))$ is 0 and that of $B(Y_i, X_i)$ is at most K [see (2.5)]. By (2.2) we have

$$|\tau_{ni}| \le \delta\pi n^{-1/2} \tag{3.9}$$

and thus

$$E_0 W_n = 0, \qquad E_0|Z_n| \le \delta^2 K \pi^2 \tag{3.10}$$

and

$$E_0 W_n^2 \le \delta^2 \pi^2 K. \tag{3.11}$$

Hence, $E_0|l_n| \le C\delta$ with C as in (3.2) and the Markov inequality and (3.3) give

$$P_0\{-\varepsilon_0 < l_n < \varepsilon_0\} \ge 1 - \varepsilon_0.$$

This implies [cf. (3.2)]

$$P_0\{1 - \varepsilon < L_n < 1 + \varepsilon\} \geq 1 - \varepsilon. \tag{3.12}$$

It is then easy to see that, for any event A,

$$P_0(A) - 2\varepsilon \leq P_1(A) \leq P_0(A) + 2\varepsilon. \tag{3.13}$$

Next choose an integer m such that

$$\frac{1}{m+1} \leq \varepsilon, \tag{3.14}$$

set $t_i = (i/m)\delta_1$ for $i = 0, \dots, m$ and use the preceding results for $\delta = t_i$, $i = 1, \dots, m$. Denote now P_1 and f_n corresponding to $\delta = t_i$ by P_i and f_{ni}.

Consider a statistical method φ based on Z_n and with the range $\{0, 1, \dots, m\}$ and with $\alpha_i = P_i\{\varphi \neq i\}$. We obtain, using (3.13) twice,

$$\alpha_i = \sum_{j \neq i} P_i\{\varphi = j\} \geq \sum_{j \neq i} (1 - \alpha_j) - 4\varepsilon$$

and thus $\sum \alpha_i \geq m - 4\varepsilon$ and

$$\alpha = \max_{i=0,\dots,m} \alpha_i \geq \frac{m}{m+1} - 4\varepsilon.$$

This and (3.2) and (3.14) give

$$\alpha \geq \eta. \tag{3.15}$$

Construct a particular φ such that $\{\varphi = i\} \supset \{|T_n - \chi(f_{ni})| \leq c\kappa_n\}$ with

$$c = t_1/3. \tag{3.16}$$

Note that $\chi(f_{ni}) = it_1 n^{-1/3}$. If (3.1) does not hold, it follows that $\alpha < \eta$, a contradiction to (3.15). This proves (3.1) for c as in (3.16). □

REMARK 3.2. In Theorem 3.1, the domain $[-1, 1]$ of the functions in F can be replaced by another bounded interval I. It can be easily changed to the infinite intervals I by choosing f in F to have restrictions to $I - [-2, 2]$ independent of f. The result can be extended to functions on I^k by choosing F such that Theorem 3.1 applies to the sections $f(\cdot, 0, \dots, 0)$ of f in F.

Acknowledgments. This work formed part of my dissertation written under the supervision of Professor C. J. Stone, whose guidance and suggestions are gratefully acknowledged. I should also like to thank a referee, who spent a considerable amount of time on the original version of this manuscript and made very helpful suggestions for improving the presentation of this paper.

REFERENCES

BOX, G. E. P. and WILSON, K. B. (1951). On the experimental attainment of optimum conditions (with discussion). *J. Roy. Statist. Soc. Ser. B* **13** 1–45.

CHEN, H. (1984). Optimal rates of convergence for locating the global maximum of a regression function. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

DEVROYE, L. P. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Trans. Inform. Theory* **IT-24** 142–151.

FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.* **38** 191–200.

FABIAN, V. and HANNAN, J. (1985). *Introduction to Probability and Mathematical Statistics*. Wiley, New York.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* **23** 462–466.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NEVELSON, I. A. and HASMINSKII, R. Z. (1973). *Stochastic Approximation and Recursive Estimation*. Amer. Math. Soc., Providence, R.I.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK 11794-3600