Lower Rate of Convergence for Locating a Maximum of a Function

Hung Chen

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 $p$th 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 $p$th 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 $\kappa_n$ cannot be improved for $F$ containing enough functions with bounded $p$th 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 $\delta_0$ 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.
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 \( n = 1, 2, \ldots \), the function
\[
(2.1) \quad f_{n\delta}(x) = -x^2 + 2\delta n^{-1/2} \arctan(n^{1/(2p)}(x - \delta \kappa_n))
\]
is in \( F \).

REMARK 2.2. Notice that
\[
(2.2) \quad |f_{n\delta} - f_{n0}| \leq \delta \pi n^{-1/2}
\]
and that
\[
(2.3) \quad \chi(f_{n0}) = 0, \quad \chi(f_{n\delta}) = \delta \kappa_n
\]
since
\[
f_{n\delta}'(x) = -2x + 2\delta \kappa_n \arctan(n^{1/(2p)}(x - \delta \kappa_n)).
\]
For all \( \delta \) sufficiently small, \( f_{n\delta}'' < -1 \) and \( |f_{n\delta}'| \leq 1 \). In particular, such \( f_{n\delta} \) are concave with a unique maximum.

CONDITION 2.3. \( \mu \) is a measure on the \( \sigma \)-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 \( \mu \). 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 \( y, x, t \),
\[
(2.4) \quad |g''(y, x, t)| \leq A(y, x), \quad |l''(y, x, t)| \leq B(y, x)
\]
with functions \( A \) and \( B \) satisfying
\[
(2.5) \quad \int A(\cdot, x) \, d\mu \leq K, \quad \int B(\cdot, x) g(\cdot, x, t) \, d\mu \leq K
\]
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 \( X_1, \ldots, X_{n-1}, Y_1, \ldots, Y_{n-1} \) does not depend on \( f \), and each \( Y_n \) has a conditional density with respect to \( \mu \), given \( X_1, \ldots, X_n, Y_1, \ldots, Y_n \), equal to \( g(\cdot, X_n, f(X_n)) \).

We shall write \( Z_n \) for \( \langle X_n, Y_n \rangle \) and \( \mathcal{Z}_n \) for \( \langle Z_1, \ldots, Z_n \rangle \).

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

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

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 \( T \),
\[
(3.1) \quad \inf_{f} P_{f,T} \{|T_n - \chi(f)| \geq c \kappa_n \} \geq \eta.
\]
PROOF. Let $n$ be a positive integer and $\eta$ a positive number smaller than 1. Set
\begin{equation}
C = \pi^2(K + \sqrt{K}), \quad \varepsilon_0 = (1 - \eta)/10, \quad \varepsilon = 2\varepsilon_0
\end{equation}
and choose $\delta_1$ positive but small enough [cf. (2.4)] that $f_{n\delta}$ has a unique maximum at $\delta_{\varepsilon_n}$ for every $n$ and every $\delta \leq \delta_1$ (cf. Remark 2.2) and such that
\begin{equation}
\delta_1 \leq \min\{\delta_0, \varepsilon_0^2/C\}.
\end{equation}
Consider an estimate $\langle \mathcal{D}, \langle T_n \rangle \rangle$ and a $\delta$ in $(0, \delta_1]$ and set
\begin{equation}
f(x) = -x^2, \quad f_n = f_{n, \delta}.
\end{equation}
Denote $P_{\mathcal{D}, f}$ by $P_0$ and $P_{\mathcal{D}, f_n}$ by $P_1$ and denote by $P_{0m}, P_{1m}$ the distribution of $\mathcal{D}_m$ under $P_0$ and $P_1$, respectively. The density of $P_{1n}$ with respect to $P_{0n}$ is
\begin{equation}
L_n = \prod_{i=1}^{n} g(Y_i, X_i, f(X_i) + \tau_{ni})/g(Y_i, X_i, f(X_i))
\end{equation}
with
\begin{equation}
\tau_{ni} = f_n(X_i) - f(X_i).
\end{equation}
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
\begin{equation}
l_n = W_n + Z_n
\end{equation}
with
\begin{equation}
W_n = \sum_{i=1}^{n} \tau_{ni}l'(Y_i, X_i, f(X_i))
\end{equation}
and
\begin{equation}
|Z_n| \leq \frac{1}{2} \sum_{i=1}^{n} \tau_{ni}^2B(Y_i, X_i).
\end{equation}
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
\begin{equation}
|\tau_{ni}| \leq \delta_\pi n^{-1/2}
\end{equation}
and thus
\begin{equation}
E_0 W_n = 0, \quad E_0 |Z_n| \leq \delta^2 K \pi^2
\end{equation}
and
\begin{equation}
E_0 W_n^2 \leq \delta^2 \pi^2 K.
\end{equation}
Hence, $E_0 |l_n| \leq C\delta$ with $C$ as in (3.2) and the Markov inequality and (3.3) give
\begin{equation}
P_0\{-\varepsilon_0 < l_n < \varepsilon_0\} \geq 1 - \varepsilon_0.
\end{equation}
This implies [cf. (3.2)]

\[(3.12) \quad P_0(1 - \varepsilon < L_n < 1 + \varepsilon) \geq 1 - \varepsilon.\]

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

\[(3.13) \quad P_0(A) - 2\varepsilon \leq P_1(A) \leq P_0(A) + 2\varepsilon.\]

Next choose an integer \(m\) such that

\[(3.14) \quad \frac{1}{m + 1} \leq \varepsilon,\]

set \(t_i = (i/m)\delta_1\) for \(i = 0, \ldots, m\) and use the preceding results for \(\delta = t_i, i = 1, \ldots, m\). Denote now \(P_i\) and \(f_n\) corresponding to \(\delta = t_i\) by \(P_i\) and \(f_n\). Consider a statistical method \(\varphi\) based on \(Z_n\) and with the range \(\{0, 1, \ldots, m\}\) 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 \(\alpha_i \geq m - 4\varepsilon\) and

\[\alpha = \max_{i=0, \ldots, m} \alpha_i \geq \frac{m}{m + 1} - 4\varepsilon.\]

This and (3.2) and (3.14) give

\[(3.15) \quad \alpha \geq \eta.\]

Construct a particular \(\varphi\) such that \(\{\varphi = i\} \supset \{\lceil T_n - \chi(f_n) \rceil \leq c\kappa_n\}\) with

\[(3.16) \quad c = t_i/3.\]

Note that \(\chi(f_n) = it_n n^{-1/3}\). 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). \(\square\)

**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 \(f(\cdot, 0, \ldots, 0)\) 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**


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