SEQUENTIAL MINIMAX SEARCH FOR A MAXIMUM

J. KIEFER[1]

The problem formulated below was motivated by that of determining an interval containing the point at which a unimodal function on the unit interval possesses a maximum, without postulating regularity conditions involving continuity, derivatives, etc. Our solution is to give, for every $\epsilon > 0$ and every specified number N of values of the argument at which the function may be observed, a procedure which is ϵ-minimax (see (1) below) among the class of all sequential nonrandomized procedures which terminate by giving an interval containing the required point, where the payoff of the computer to nature is the length of this final interval. (The same result holds if, e.g., we consider *all* nonrandomized procedures and let the payoff be length of interval plus c or 0 according to whether the interval fails to cover or covers the desired point, where $c \geq 1/U_{N+1}$, the latter being defined below.) The analogous problem where errors are present in the observations was considered in [1], but no optimum results are yet known for that more difficult case.

Search for a maximum is a "second-order" search in the sense that information is given by *pairs* of observations. Thus, if $x_1 < x_2$ and $f(x_1) \geq f(x_2)$, and f is a member of the class $\mathcal{J}$ described below, the point $x^{(f)}$ defined below (the point where f attains its maximum if the latter exists) must lie to the left of x_2. If we postulate only that $f \in \mathcal{J}$, this is essentially the only information we have about $x^{(f)}$: i.e., that future observations should be taken to the left of x_2. Similarly, the problem of finding the point $x^{[f]}$ at which a strictly increasing function f on the unit interval attains a value α (weaker restrictions on f can be made) is essentially first-order in the sense that every single observation gives information about where to take the next. (See [2; 3] for the corresponding statistical problem.) The minimax procedure in that case is successively to split the interval in which $x^{[f]}$ is known to lie into equal parts at the point of next observation, in an obvious manner. It may be fruitful to consider higher-order search problems, as will be done in a future paper.

Let the class $\mathcal{J}$ consist of every function f from the closed unit interval I into the reals (R) and for which there is an $x^{(f)} \in I$ such that f is either strictly increasing for $x \leq x^{(f)}$ and strictly decreasing for $x > x^{(f)}$, or else strictly increasing for $x < x^{(f)}$ and strictly decreasing

for $x \geq x^{(f)}$. Let $\mathcal{D}$ denote the space of all closed intervals D which are subsets of I. An element $D \in \mathcal{D}$ will be called a terminal decision. Let N be given integer ≥ 2. A (nonrandomized) strategy S is a set $S = \{x_1, g_2, \cdots, g_N, s, t\}$ consisting of a number $x_1 \in I$, functions g_k from $I^{k-2} \times R^{k-1}$ into I $(2 \leq k \leq N)$, and functions s and t from $I^{N-1} \times R^N$ into I and with $s \leq t$. A strategy S is used as follows: one observes (or computes) in order

$$f(x_1), f(x_2), \cdots, f(x_N) \text{ with } x_k = g_k[x_2, \cdots, x_{k-1}, f(x_1), \cdots, f(x_{k-1})]$$

for $2 \leq k \leq N$ $(x_2 = g_2[f(x_1)])$, and then selects the closed interval

$$D(f, S) = [s(x_2, \cdots, x_N, f(x_1), \cdots, f(x_N)),$$
$$t(x_2, \cdots, x_N, f(x_1), \cdots, f(x_N))]$$

as terminal decision. Let $\mathcal{S}_N$ be the class consisting of every strategy S requiring N observations and for which $x^{(f)} \in D(f, S)$ for all f in $\mathcal{F}$. Our problem is, given any $\epsilon > 0$ and $N \geq 2$, to find an $S_N^* \in \mathcal{S}_N$ such that

$$(1) \qquad \sup_{f \in \mathcal{F}} L(D(f, S_N^*)) \leq \inf_{S \in \mathcal{S}_N} \sup_{f \in \mathcal{F}} L(D(f, S)) + \epsilon,$$

where $L(D)$ is the length of the interval D.

We shall now describe an S_N^* which, it will be shown, satisfies (1). Let U_n be the nth Fibonacci number $(U_0 = 0, U_1 = 1, U_n = U_{n-1} + U_{n-2}$ for $n \geq 2)$. S_2^* is defined by $x_1 = 1/2$, $x_2 = 1/2 + \epsilon$, and $[s, t] = [0, x_2]$ or $[x_1, 1]$ according to whether $f(x_1) \geq f(x_2)$ or $f(x_1) < f(x_2)$. Suppose S_{N-1}^* has been defined $(N \geq 3)$. We then define S_N^* as follows: $x_1 = U_{N-1}/U_{N+1}$, $x_2 = 1 - x_1 = U_N/U_{N+1}$. According to whether $f(x_1) \geq f(x_2)$ or $f(x_1) < f(x_2)$, let $h(x) = xU_{N+1}/U_N$ or $h(x) = (-U_{N-1} + xU_{N+1})/U_N$, let $y = h(x)$ and $f^*(y) = f(h^{-1}(y))$ for $y \in I$, and define $y_2 = h(x_1)$ or $y_1 = h(x_2)$ in the respective cases. Thus, $y_2 = U_{N-1}/U_N$ or $y_1 = U_{N-2}/U_N$. Use S_{N-1}^* on the variable y and function f^* (for $y \in I$), noting that either $f^*(y_2)$ or $f^*(y_1)$ has already been observed. Obviously, $S_N^* \in \mathcal{S}_N$.

We shall now prove inductively that S_N^* satisfies (1). This is obvious for $N = 2$. Assuming it to be true for $N \leq n$, we now prove (1) to hold for $N = n+1$. Suppose the latter to be false. Then, since $L(D(f, S_N^*)) \leq \epsilon + 1/U_{N+1}$, there would exist a procedure $\bar{S} \in \mathcal{S}_{n+1}$ for which

$$(2) \qquad \sup_{f \in \mathcal{F}} L(D(f, \bar{S})) < 1/U_{n+2}.$$

We may suppose that, under $\bar{S}$, g_2 is a constant. For otherwise we

could define a procedure $\widehat{S}$ by using the x_1 of $\overline{S}$ and thereafter using $\overline{S}$ on the function $f(x)-f(x_1)$; the g_2 of this procedure is then constant (equal to the $g_2(0)$ of $\overline{S}$), and $\widehat{S}$ clearly satisfies (2) (with $\overline{S}$ replaced by $\widehat{S}$) if $\overline{S}$ does. We hereafter denote the x_1 and x_2 of $\overline{S}$ by b and $b+a=1-c$ (say) with $a\geq 0$; the corresponding values for S_{n+1}^* will be denoted by $d=U_n/U_{n+2}$ and $d+e=1-d=U_{n+1}/U_{n+2}$.

We next show that, as a consequence of (2),

$$(3) \qquad\qquad a+b \leq d+e \quad\text{and}\quad a+c \leq d+e.$$

We prove the first inequality of (3), the proof of the second being similar. Suppose, to the contrary, that $a+b>d+e$. We shall construct a procedure $S'\in\mathcal{S}_n$ for which

$$(4) \qquad\qquad \sup_{f\in\mathcal{J}} L(D(f,S')) < 1/U_{n+1},$$

contradicting the induction hypothesis. For any $f\in\mathcal{J}$ define f' for $y\in I$ by

$$(5) \qquad f'(y) = \begin{cases} \exp\left\{f\left(\dfrac{y}{a+b}\right)\right\} & \text{if}\quad 0\leq y < a+b, \\[2mm] -y & \text{if}\quad a+b\leq y\leq 1. \end{cases}$$

It is easy to verify that $f'\in\mathcal{J}$ and that $x^{(f')}=(a+b)x^{(f)}$. Under S' we observe $\exp\ f(b/(a+b))=f'(b)$, calculate $f'(a+b)=-a-b$, treat these as the first two observations on f' according to $\overline{S}$, and take the remaining $n-1$ observations on f under S' by using $\overline{S}$ on f' as follows: if the kth observation on f' under $\overline{S}$ is to be taken at a value $y_k\geq a+b$ (for $k\geq 3$), S' puts $x_{k-1}=0$ and ignores the value of $f(x_{k-1})$, $f'(y_k)$ is calculated from the last line of (5), and y_{k+1} is computed under $\overline{S}$; if the kth observation on f' under $\overline{S}$ is to be taken at a value $y_k<a+b$, the $(k-1)$th observation on f according to S' is taken at $x_{k-1}=y_k/(a+b)$, the value of $f'(y_k)$ is then computed from (5) from the observed $f(x_{k-1})$, and y_{k+1} is determined by $\overline{S}$. After n observations on f ($n+1$ on f'), we put $D(f,\ S')=[s'/(a+b),$ $\min\ (1,\ t'/(a+b))]$, where $[s',\ t']=D(f',\ \overline{S})$. Clearly $x^{(f)}\in D(f,S')$, so that $S'\in\mathcal{S}_N$. Moreover, since $L(D(f,\ S'))\leq L(D(f',\ \overline{S}))/(a+b)$, if $a+b>d+e=U_{n+1}/U_{n+2}$, equation (2) would imply equation (4). This completes the proof of (3).

The second inequality of (3) and the fact that $a+b+c=2d+e=1$ show that $b\geq d$ (similarly, $c\geq d$). We shall use this and (2) to construct a procedure $S''\in\mathcal{S}_{n-1}$ for which

$$(6) \qquad\qquad \sup_{f\in\mathcal{F}} L(D(f, S'')) < 1/U_n;$$

this contradiction of our induction hypothesis will then imply that (2) is false, completing our proof. To this end, for any $f\in\mathcal{F}$ define f'' for $y\in I$ by

$$(7) \qquad f''(y) = \begin{cases} \exp\left\{f\left(\dfrac{y}{b}\right)\right\} & \text{for } 0 \leq y < b, \\[2mm] -y & \text{for } b \leq y \leq 1. \end{cases}$$

Clearly, $f''\in\mathcal{F}$ and $x^{(f'')}=bx^{(f)}$. S'' is defined by using $\bar{S}$ on f'' in a manner similar to that used in the previous paragraph on f' to define S': the first two observations on f'' under $\bar{S}$ are $f''(b)=-b$ and $f''(a+b)=-a-b$; thereafter, the kth observation on f'' under $\bar{S}$ corresponds in an obvious manner to the $(k-2)$th on f under S'', and $D(f, S'')=[s''/b,\ \min\ (1,\ t''/b)]$, where $[s'',\ t'']=D(f'',\ \bar{S})$. Thus, $L(D(f,\ S''))\leq L(D(f'',\ \bar{S}))/b$; since $b\geq d=U_n/U_{n+2}$, this and (2) imply (6), completing the proof of our assertion.

We remark that a minimax procedure (one satisfying (1) with $\epsilon=0$) does not exist in the above problem, as is evident from the case $N=2$. The procedure S_N^* defined above may be improved upon by noting that, whenever the two largest observations are equal, $x^{(f)}$ must lie between the two corresponding values of x for any $f\in\mathcal{F}$. An interesting procedure which is not minimax for any fixed N but will often be useful in applications is the strategy S^* defined as follows: let $x_2=1-x_1=-1/2+5^{1/2}/2=.618=\mu$ (say). If $f(x_1)\geq f(x_2)$, define $v(x) =x/\mu$ and $y=v(x)$, and $f^*(y)=f(v^{-1}(y))$ for $y\in I$. Putting $y_2=v(x_1)=\mu$ and $y_1=1-\mu$, we then use S^* on the variable y and function f^*, where we already have observed $f^*(y_2)$. (A similar procedure applies if $f(x_1)<f(x_2)$.) Continuing in this manner, at every stage we have the same geometric configuration, unlike the case of S_N^*. The advantage of this is that if the number of observations is not specified in advance but is determined after several values have been observed (e.g., more observation might be taken if f appears to be sharply peaked near its maximum), the use of any S_N^* (or sequence of S_N^*'s) can lead to great inefficiency if one decides after N observations to take more. When N is large, if S^* is used for N observations, the length of the final interval is about 1.17 times that of S_N^* (with $\epsilon\to0$).

REFERENCES

1. J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, Ann. Math. Statist. vol. 23 (1952) pp. 462–466.

2. H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statist. vol. 22 (1951) pp. 400–407.

3. J. Wolfowitz, *On the stochastic approximation method of Robbins and Monro*, Ann. Math. Statist. vol. 23 (1952) pp. 457–461.

CORNELL UNIVERSITY

MEAN VALUE METHODS IN ITERATION

W. ROBERT MANN

Due largely to the works of Cesàro, Fejér, and Toeplitz, mean value methods have become famous in the summation of divergent series. The purpose of this paper is to show that the same methods can play a somewhat analogous role in the theory of divergent iteration processes. We shall consider iteration from the limited but nevertheless important point of view of an applied mathematician trying to use a method of successive approximations on some boundary value problem which may be either linear or nonlinear.

It is now widely known that the Schauder fixpoint theorem [1] is a powerful method for proving existence theorems. If one wishes to use it to prove that a given problem has a solution, he proceeds by associating with the problem a convex compact set E in some Banach space, and a continuous transformation T which carries E into itself. Schauder's theorem asserts that T must have at least one fixpoint, say p, in E. If E and T have been appropriately chosen, it can then usually be shown that any such fixpoint must be a solution of the original problem and conversely. Mathematical literature since about 1935 abounds with illustrations of this technique. We mention here only [2] and [3] which contain the genesis of the present work.

Let us then begin with a convex compact set E in a Banach space, and a continuous transformation T carrying E into itself. The problem which we shall consider is that of constructing in E a sequence of elements $\{x_n\}$ that converge to a fixpoint of T. Ordinarily one starts by choosing more or less arbitrarily an initial point x_1 in E and then considering the successive iterates $\{x_n\}$ of x_1 under T, where

$$(1) \qquad\qquad x_{n+1} = T(x_n).$$

If this sequence converges, then obviously its limit point is a fixpoint of T and the problem is solved. But to guarantee convergence one must impose some further restriction on T, such as, for example, that

---