THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Department of Communication Sciences

Technical Report

A CLASS OF SEQUENTIAL SAMPLING PROBLEMS ARISING IN CERTAIN LEARNING SITUATIONS

Edwin Bainbridge, E.S.

ORA PROJECTS 08226, 03105 and 01252

supported by:

U.S. ARMY RESEARCH OFFICE (DURHAM)
CONTRACT NO. DA-31-124-ARO-D-483
DURHAM, NORTH CAROLINA

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
CONTRACT NO. Nonr-1224(21)
WASHINGTON D.C.

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE
NATIONAL INSTITUTES OF HEALTH
GRANT NO. GM-12236-04
BETHESDA, MARYLAND

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

December 1967

RESEARCH PROGRESS REPORT

Title: "A Class of Sequential Sampling Problems Arising in Certain Learning Situations," E. S. Bainbridge, University of Michigan Technical Report 03105-49-T.

Background: The Logic of Computers Group of the Communication Sciences Department of The University of Michigan is investigating the application of logic and mathematics to the design of computing automata. The application of the techniques and concepts of mathematics, including sequential sampling techniques, to the study of automaton learning, forms a part of this investigation.

Condensed Report Contents: A strategist is to decide on each of n turns whether to take a sample from a certain fixed random variable, and receive the outcome as payoff, or to receive as payoff the largest value of the random variable he has discovered so far. The expected total payoff for the n turns is to be maximized. It is shown that the following decision procedure is the solution.

$$\text{Sample until } \frac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{\hat{x}}^{\infty}(1-F(x))dx} \text{ exceeds the number of turns remaining}$$

where $\hat{x}$ is the current record value, and $F(x)$ is the distribution function of the random variable.

A strategy for an indefinite number of turns is described, and for suitable distributions it is shown that the limit of the ratio of the payoff accumulated by this strategy in n turns to the payoff accumulated by the optimal strategy for n turns is one, with probability one.

For Further Information: The complete report is available in the major Navy technical libraries and can be obtained from the Defense Documentation Center. A few copies are available for distribution by the author.

Introduction

In a competitive learning situation, the individual must not only discover strategies which yield high payoffs, but also spend a certain part of his time actually using these strategies in order to accumulate enough payoff to maintain his competitive position. Typically, the type of activity which produces information which may be used to improve his strategy is quite distinct from the activity involved in the use of the current best strategy, and will yield considerably lower payoffs, if any. For example, market research presumably produces a better advertising campaign (from the point of view of the manufacturer), but the money spent there increases revenue only when the new campaign is begun. It is therefore necessary for the individual with limited resources to make decisions between information producing and payoff producing activities. In this example, the individual may, if he has sufficient resources, elect to devote some to the gathering of information, and some to the gathering of payoff. However, the resources may be so restricted that only one of these activities may be undertaken at any time. For example, a chess player must decide whether to try an opening with which he is relatively unfamiliar in case it should prove more effective than the one he typically uses, or to stick with his familiar favorite opening. The situation we examine in this paper is inspired by the latter example.

A special case of this type of learning situation may be termed trial and error learning. We imagine the player to have available a class of strategies, some of which have been tried, and others which have not. On each turn the player is to decide whether to try a new one, or to use the strategy which has been found to provide the highest payoff. In any real situation, if the player decides to attempt a new strategy, he will choose

one which he thinks has a reasonable chance of success. Also, even if the strategy he tries does not prove better than the current best strategy, the player will presumably get information which he may use in deciding which untried strategy he will attempt next. It would therefore seem unrealistic to suppose that when the player chooses a new strategy that he does so at random. However, we may redefine the problem as follows. We may consider the strategy by which the player analyzes unsuccessful attempts in order to decide what his next trial shall be. If he always assimilates information from his explorations in the same way, then surely there is an advantage to be gained in sometimes assimilating it in a different way. If he does change it from time to time, then he has exactly the same problem with these second level strategies as he does with his first level strategies. We therefore argue that unless he has an infinite heirarchy of strategies, at some level either he selects new strategies randomly, or it is to his advantage to do so. It is this problem, in which new strategies are selected randomly, and the only information gained from them is whether or not they were an improvement on the currently best strategy, that we call trial and error learning, and discuss in this paper.

This then is the motivation for our formulation, which is the following. Each turn must be either a play of the best strategy, which is assumed to produce the same payoff each time it is used; or the turn consists of a trial of a new strategy, which is selected randomly, and hence produces a payoff from some fixed probability distribution. We first consider the case in which a fixed number of turns are allotted, and the expected total payoff for those turns is to be maximized. The solution is explicitly given for the case of a continuous distribution with a mean, and involves a certain criterion function. This criterion function is then used to define

a strategy which does not depend on any fixed number of turns. For suitable distributions, this strategy is shown with probability one to have a payoff ratio to the class of best strategies for fixed numbers of turns which approaches one. It is also pointed out that if, as one would expect to be the case, the payoff distribution is not known, then the form of the criterion function is such that it may be estimated as play proceeds.

I. The Basic Decision Problem

Let x be a fixed random variable. We consider a class of decision problems, each of which is characterized by a positive integer n and a real number $\hat{x}$. The decision problem $(n,\hat{x})$ is given as follows. On each of n turns, a decision is to be made between two alternatives, which we shall call sampling and collecting. The decision to sample results in the strategist receiving as payoff the result of an independent sample from the random variable x. The decision to collect results in the strategist receiving as payoff the current record value, where the record value on each turn of the decision process is defined as follows. On the first turn, the record value is $\hat{x}$. On successive turns, the record value is the larger of the record value on the previous turn, and the payoff received on the previous turn. The strategist is to maximize the expected total payoff for the n turns.

We obtain the decision procedure which does this as follows. Let us assume that the random variable x has a mean $\bar{x}$, and write $\phi_n(\hat{x})$ for the maximum expected total payoff over all decision procedures for the problem $(n,\hat{x})$. If a strategist samples on the first turn, he receives the payoff x, and his new record value is $\max(x,\hat{x})$. If he collects on the first turn, he receives the payoff $\hat{x}$, and his record value remains $\hat{x}$. Thus we have the recursion

$$\phi_n(\hat{x}) = \max \begin{cases} \bar{x} + E(\phi_{n-1}(\max(x,\hat{x}))) \\ \hat{x} + \phi_{n-1}(\hat{x}) \end{cases} \tag{1}$$

Since $\phi_1(\hat{x}) = \max(\bar{x},\hat{x})$, we may calculate $\phi_n(\hat{x})$ for all $(n,\hat{x})$ if we know the distribution of x.

Furthermore, let us define $\psi_n(\hat{x}) = \hat{x} + \phi_{n-1}(\hat{x}) - \bar{x} - E(\phi_{n-1}(\max(x,\hat{x})))$. Then if r turns remain, and the current record value is $\hat{x}$, the optimal decision for the next turn is

$$\begin{aligned} &\text{sample if } \psi_r(\hat{x}) < 0 \\ &\text{collect if } \psi_r(\hat{x}) \geq 0 \end{aligned} \tag{2}$$

since by making this decision the strategist chooses the alternative leading to the larger potential total expected value.

Let us further assume that x has a continuous density function f(x), with distribution function F(x). Then the following facts are readily verified by induction.

(i) $\phi_n(\hat{x})$ is continuous and piecewise differentiable for each n, with $\phi_n'(\hat{x}) \geq 0$ when defined.

(ii) $\psi_n(\hat{x})$ is continuous and piecewise differentiable for each n, with $\psi_n'(\hat{x}) \geq 1$.

(iii) $\psi_n(\hat{x}) \leq \hat{x} - \overline{x}$, and hence is negative for $\hat{x} < \overline{x}$.

If x is unbounded, then $\psi_n(\hat{x})$ is ultimately positive by (ii), and if x is bounded, say by x_{max}, then $\psi_n(x_{max}) > 0$. Thus in either case, $\psi_n(x)$ has positive values, and by (iii), negative values, and by (ii) is strictly increasing, and hence has a unique zero, which we denote by z_n. We may thus restate the optimal decision rule when r turns remain and the current record value is $\hat{x}$, as

$$\begin{aligned} &\text{sample if } \hat{x} < z_r \\ &\text{collect if } \hat{x} \geq z_r \end{aligned} \tag{3}$$

We now verify that

$$z_{n+1} \geq z_n \tag{4}$$

First, $\psi_n(z_n) = 0$ by definition, so that

$$z_n + \phi_{n-1}(z_n) = \overline{x} + E(\phi_{n-1}(\max(x,z_n))) \ .$$

Thus $\phi_n(z_n) = z_n + \phi_{n-1}(z_n) = \overline{x} + E(\phi_{n-1}(\max(x,z_n)))$, by (1). Now we see that $\psi_{n+1}(z_n)$ is not positive, since

$$\begin{aligned} \phi_{n+1}(z_n) &= z_n + \phi_n(z_n) - \overline{x} - E(\phi_n(\max(x,z_n))) \\ &= z_n + [\overline{x} + E(\phi_{n-1}(\max(x,z_n)))] - \overline{x} - E(\phi_n(\max(x,z_n))) \end{aligned}$$

and noting that for any y, $\phi_n(y) - \phi_{n-1}(y) \geq y$, then taking $y = \max(x,z_n)$,

we obtain $\psi_{n+1}(z_n) \leq z_n - E(\max(x,z_n)) \leq 0$. Thus we must have $z_n \leq z_{n+1}$, as claimed, since $\psi_{n+1}(\hat{x})$ is increasing in x.

Now suppose $\hat{x} \geq z_n$. Then $\psi_n(\hat{x}) \geq \psi_n(z_n) = 0$, so that

$$\hat{x} + \phi_{n-1}(\hat{x}) \geq \overline{x} + E(\phi_{n-1}(\max(x,\hat{x})))$$

and so $\phi_n(\hat{x}) = \hat{x} + \phi_{n-1}(\hat{x})$, by (1). But $\hat{x} \geq z_n \geq z_{n-1}$, hence we obtain by induction that

$$\phi_n(\hat{x}) = nx \text{ for } \hat{x} \geq z_n \tag{5}$$

We are now in a position to obtain the optimal decision rule in an explicit form. Since $\psi_n(z_n) = 0$ and $\phi_{n-1}(z_n) = (n-1)z_n$, $\phi_{n-1}(\max(x,z_n)) = (n-1)\max(x,z_n)$, we obtain by substitution that

$$z_n + (n-1)z_n = \overline{x} + (n-1)E(\max(x,z_n)), \text{ and hence}$$

$$\frac{z_n - \overline{x}}{n-1} = E(\max(x,z_n)) - z_n \tag{6}$$

Now for any continuous distribution with a mean the following facts are easily verified by writing, e.g., $1 - F(x) = \int_x^\infty f(y)dy$ and interchanging the order of integration.

(iv) $E(\max(x,x_o)) - x_o = \int_{x_o}^\infty (1-F(x))dx$

(v) $x_o - \overline{x} = \int_{-\infty}^{x_o} F(x)dx - \int_{x_o}^\infty (1-F(x))dx$.

Thus taking $x_o = z_n$, substituting in (6) and simplifying, we obtain

$$\frac{\int_{-\infty}^{z_n} F(x)dx}{\int_{z_n}^\infty (1-F(x))dx} = n \ . \tag{7}$$

Let us now define

$$z(\hat{x}) = \frac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{\hat{x}}^\infty (1-F(x))dx} \tag{8}$$

We note that $z(\hat{x})$ is strictly increasing, and $z(z_n) = n$ by (7) so that we may once again restate our optimal decision rule when r turns remain and

the current record value is $\hat{x}$, as

$$\begin{aligned} &\text{sample if } z(\hat{x}) < r \\ &\text{collect if } z(\hat{x}) \geq r \end{aligned} \tag{9}$$

Furthermore, the value $z(\hat{x})$ where $\hat{x}$ is the current record value, may be estimated on the basis of the observations on x taken to that point. Let us rewrite $z(\hat{x})$ using (v), as

$$\frac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{-\infty}^{\hat{x}} F(x)dx - \hat{x} + \bar{x}} \quad .$$

We note that in this form the portion of the distribution of x about which no evidence has been received, that is, the part above $\hat{x}$, does not appear explicitly.

In summary, we have shown that if x is a random variable with mean $\bar{x}$, continuous density function f(x) and distribution function F(x), then the decision which maximizes expected future total payoff when r turns remain and the current record value is $\hat{x}$, is

$$\begin{aligned} &\text{sample if } z(\hat{x}) < r \\ &\text{collect if } z(\hat{x}) \leq r \end{aligned}$$

where $z(\hat{x}) = \dfrac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{\hat{x}}^{\infty}(1-F(x))dx} = \dfrac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{-\infty}^{\hat{x}} F(x)dx - \hat{x} + \bar{x}}$

is a function which may be estimated from the samples so far taken.

II. The Extended Decision Problem

It is natural to ask whether there is an acceptable strategy if the number of turns available is not fixed. It is not clear what it would mean for such a strategy to be optimal, since nay such strategy can be improved in the following sense. Let X be the sample space consisting of infinite sequences $\vec{x} = (x_1, x_2, \ldots)$ of independent samples from the random variable x. A strategy for an indefinite number of turns is characterized by a function S mapping X into the set Δ of infinite sequences $\vec{\delta} = (\delta_1, \delta_2, \ldots)$ of binary digits δ_r. The interpretation of such a sequence is that a sample is taken on the r^{th} turn just in case $\delta_r = 1$. We write the r^{th} component of S as S_r. We have perhaps made this definition too broad since not all such strategies are realizable in the sense that the decision at turn r may depend on samples not yet taken. However, this does not matter, since we shall construct from S an improved strategy S* which is realizable if S is. We first exclude those strategies which after some point either never sample, or never collect, that is, those strategies for which

$$P(\exists n\ \forall r\ r > n \Longrightarrow S_r = 0) = 1$$

or

$$P(\exists n\ \forall r\ r > n \Longrightarrow S_r = 1) = 1$$

The former type can be improved in an obvious way by sampling beyond the point at which sampling would have stopped, until a better record value is found, and then collecting indefinitely; and the latter can be improved by collecting indefinitely once a record value better than the mean has been found. Let us consider the more interesting cases more explicitly. If S is not of the two previous types, then for almost all $\vec{x} \in X$, play proceeds by alternate blocks of sampling and collecting. We can therefore pick some fixed n such that with positive probability at least two blocks of collecting separated by a block of sampling have occurred in n turns. We define

a strategy S* by

$$S_r^* = S_r \quad \text{for} \quad r > n$$

and for $r \leq n$

$$S_r^* = \begin{cases} 1 \text{ if } r \leq \sum_{k=1}^{n} S_k \\ 0 \text{ if } r > \sum_{k=1}^{n} S_k \end{cases}$$

That is, we have arranged that in the first n turns, all sampling is done before any collecting. If two blocks of collecting separated by a block of sampling occurred in the first n turns, then with positive probability a better record value was found before the second block began. Now since the strategy S* does all collecting in the latter part of the first n turns, it will have this better record value available when it begins collecting, and will thus do better than S, which made some collections with the smaller record value. Thus with positive probability, S* accumulates more payoff in the first n turns than S does.

Furthermore, if S was realizable, then S* was realizable, as follows. S* can be effected by using S as a "subroutine", and sampling whenever S calls for a sample, but ignoring calls for a collection until the number of samples plus the number of collections S has called for equals n. At this point the number of turns actually taken by S* is only the number of samples that S called for, and the remainder of the n turns are still available to S* for collection. Note that at each stage, say when s samples and k calls for collection have been made ($s + k \leq n$), the course of play for the first $s + k$ turns if S had been used can be reconstructed. The samples were actually taken and the outcome of each collection S called for would have been the record value at that time. Thus it is actually possible to use S in

the manner described, since any information on which S's decisions might be conditional is available to the strategist who is using S*.

This method of improving an arbitrary strategy S was chosen instead of the more straightforward method of playing the optimal strategy for the first n turns and then reverting to S, since it illustrates the following fact. Each strategy S induces a family $\{S^{(n)}\}$ of strategies for n turns which are essentially S modified so as to take advantage of the knowledge that only n turns are available. $S^{(n)}$ is given by

$$S_r^{(n)} \begin{cases} 1 \text{ if } r \leq \sum_{k=1}^{n} S_k \\ 0 \text{ if } r > \sum_{k=1}^{n} S_k \end{cases} \tag{10}$$

for $r \leq n$. We cannot find a strategy which is optimal in a strong sense, but we can find a strategy $\overline{S}$ whose induced finite strategies $\{\overline{S}^{(n)}\}$ are the optimal strategies for n turns. Furthermore under suitable assumptions the total payoff received in n turns by $\overline{S}$ is assymptotically that received by $\overline{S}^{(n)}$ for almost all $\vec{x} \in X$.

We specify $\overline{S}$ by the decision rule:

When r collections have been made, and the current record value is $\hat{x}$,

$$\begin{aligned} &\text{sample if } z(\hat{x}) < r + 1 \\ &\text{collect if } z(\hat{x}) \geq r + 1 \end{aligned} \tag{11}$$

We now establish that the induced strategies $\{\overline{S}^{(n)}\}$ are indeed the optimal strategies for n turns. Suppose the initial record value is $\hat{x}$, and consider an arbitrary $\vec{x} = (x_1, x_2, \ldots) \in X$. We know the optimal strategies do all sampling before any collecting, so it is sufficient to show that for each n, $\vec{x}$, the optimal strategy for n turns takes $\sum_{k=1}^{n} \overline{S}_r(\vec{x})$ samples.

We proceed by induction. The optimal strategy for one turn samples just in case $z(\hat{x}) < 1$ and so takes $\overline{S}_1(\vec{x})$ samples.

Let us assume that the optimal strategy for n turns takes $\sum_{r=1}^{n} \overline{S}_r(\vec{x}) = s$, say, samples in n turns. Then since the $s+1^{th}$ sample was not taken, i.e., when n - s turns remained and the current record value w as $\max(\hat{x},x_1,\ldots,x_s)$ the decision was made not to sample, we have by (9) that

$$z(\max(\hat{x},x_1,\ldots,x_s)) \geq n - s \ .$$

We examine two cases

(a) $z(\max(\hat{x},x_1,\ldots,x_s)) < n - s + 1$.

If this is the case, then $\overline{S}_{n+1}(\vec{x}) = 1$, since by our induction hypothesis $\overline{S}$ has taken s samples in n turns, and so n - s collections have been made, so that (a) together with (10) imply that $\overline{S}$ will sample on the $n+1^{th}$ turn. However, after s turns the optimal strategy for n+1 turns has made s samples [it is immediately verified independently by induction that for each $\vec{x} \in X$ the optimal strategy for n+1 turns samples at least as many times as the optimal strategy for n turns] and so has (n+1) - s turns remaining, and a record value of $\max(\hat{x},x_1,\ldots,x_s)$, so that (a) together with (9) imply that the optimal strategy for n+1 turns takes at least s+1 samples. However

$$z(\max(\hat{x},x_1,\ldots,x_{s+1})) \geq z(\max(\hat{x},x_1,\ldots,x_s)) \geq n - s = (n+1) - (s+1)$$

so that the optimal strategy for n+1 turns does not sample on the $(s+2)^{th}$ turn, and hence takes exactly

$$s + 1 = \sum_{r=1}^{n} \overline{S}_r(\vec{x}) + \overline{S}_{n+1}(\vec{x}) = \sum_{r=1}^{n+1} \overline{S}_r(\vec{x})$$

samples.

(b) $z(\max(\hat{x},x_1,\ldots,x_s)) \geq n - s + 1$.

If this is the case, then $\overline{S}_{n+1}(\vec{x}) = 0$, by (10). But also, by (9) the optimal strategy for n+1 turns will not sample on the $s+1^{th}$ turn, and hence

takes exactly

$$s = \sum_{r=1}^{n} \overline{S}_r(\vec{x}) + \overline{S}_{n+1}(\vec{x}) = \sum_{r=1}^{n+1} \overline{S}_r(\vec{x})$$

samples. We have thus established that $\overline{S}^{(n)}$ is the optimal strategy for n turns.

We now investigate the assymptotic properties of the payoff accumulated by $\overline{S}$.

We define the random variable s_n by

$$s_n(\vec{x}) = \sum_{r=1}^{n} S_r(\vec{x}) \tag{12}$$

and, if the initial record value is $\hat{x}$, we define the random variable $\hat{x}_n$ by

$$\hat{x}_n(\vec{x}) = \max(\hat{x}, x_1, \ldots, x_{s_n(\vec{x})}) \tag{13}$$

We define p_n to be the total payoff received by $\overline{S}$ in n turns, so that

$$p_n(\vec{x}) = \sum_{r=1}^{s_n(\vec{x})} x_r + \sum_{r=1}^{n} (1-\overline{S}_r(\vec{x}))\hat{x}_{r-1}(\vec{x}) \tag{14}$$

and we define $p^{(n)}$ to be the total payoff received by $\overline{S}^{(n)}$ in n turns, so that

$$p_n(\vec{x}) = \sum_{r=1}^{s_n(\vec{x})} x_r + (n-s_n(\vec{x}))\hat{x}_{s_n(\vec{x})}(\vec{x})$$

However since $\overline{S}^{(n)}$ takes no samples after turn s_n, then $\hat{x}_{s_n(\vec{x})}(\vec{x}) = \hat{x}_n(\vec{x})$, so that

$$p^{(n)}(\vec{x}) = \sum_{r=1}^{s_n(\vec{x})} x_r + (n-s_n(\vec{x}))\hat{x}_n(\vec{x}) \tag{15}$$

Thus, noting that $1 - \overline{S}_r = r - s_r - ((r-1) - s_{r-1})$, we have by summation by parts that

$$\sum_{r=1}^{n} (r-s_r)(\hat{x}_r - \hat{x}_{r-1}) = (n-s_n)\hat{x}_n - \sum_{r=1}^{n} (1-\overline{S}_r)\hat{x}_{r-1}$$

so that

$$p^{(n)} - p_n = \sum_{r=1}^{n} (r-s_r)(\hat{x}_r - \hat{x}_{r-1}) \tag{16}$$

For any $r \leq n$, let us define $\bar{r}$ to be the greatest non-negative integer less than r such that a sample was taken on turn $\bar{r} + 1$, but not on turn $\bar{r}$, if such exists; and $\bar{r} = r$ otherwise. If a sample was taken on turn r, then $\bar{r} \neq r$ and we have, since a sample was not taken on turn $\bar{r}$, that

$$z(\hat{x}_{\bar{r}-1}) \geq \bar{r} - 1 - s_{\bar{r}-1} + 1 ,$$

by (11) and

$$\hat{x}_{\bar{r}-1} = \hat{x}_{\bar{r}}$$

and

$$s_{\bar{r}-1} = s_{\bar{r}} .$$

Now a sample was taken on turn $\bar{r} + 1$, so $s_{\bar{r}+1} = s_{\bar{r}} + 1$. Thus, substituting, we obtain

$$z(\hat{x}_{\bar{r}}) \geq (\bar{r}+1) - s_{\bar{r}+1} .$$

Furthermore, since samples were taken on every turn from $\bar{r} + 1$ to r we have $(\bar{r}+1) - s_{\bar{r}+1} = r - s_r$. Thus we have

$$S_r = 1 \Longrightarrow z(\hat{x}_{\bar{r}}) \geq r - s_r . \tag{17}$$

Now $(\hat{x}_r - \hat{x}_{r-1}) = 0$ except possibly for those r such that a sample was taken on turn r. Thus we have

$$p^{(n)} - p_n = \sum_{r=1}^{n} (r-s_r)(\hat{x}_r - \hat{x}_{r-1}) \leq \sum_{r=1}^{n} z(\hat{x}_{\bar{r}})(\hat{x}_r - \hat{x}_{r-1}) .$$

We now note that if a sample was taken on turn r

$$\bar{r} \leq r - 1, \text{ so } z(\hat{x}_{\bar{r}}) \leq z(\hat{x}_{r-1}).$$

Also if r is between $\bar{n}$ and n, then $\bar{r} = \bar{n}$, and $\hat{x}_{\bar{r}} = \hat{x}_{\bar{n}}$. Thus we have

$$p^{(n)} - p_n \leq \sum_{r=1}^{\bar{n}} z(\hat{x}_{r-1})(\hat{x}_r - \hat{x}_{r-1}) + \sum_{r=\bar{n}+1}^{n} z(\hat{x}_{\bar{n}})(\hat{x}_r - \hat{x}_{r-1})$$

$$= \sum_{r=1}^{\bar{n}} z(\hat{x}_{r-1})(\hat{x}_r - \hat{x}_{r-1}) + z(\hat{x}_{\bar{n}})(\hat{x}_n - \hat{x}_{\bar{n}})$$

Now it is clear that

$$\sum_{r=1}^{\bar{n}} z(\hat{x}_{r-1})(\hat{x}_r - \hat{x}_{r-1}) \leq \int_{\hat{x}}^{\hat{x}_{\bar{n}}} z(x)dx$$

since z is increasing.

Thus

$$p^{(n)} - p_n \leq \int_{\hat{x}}^{\hat{x}_{\bar{n}}} z(x)dx + z(\hat{x}_{\bar{n}})(\hat{x}_n - \hat{x}_{\bar{n}}) \tag{18}$$

Furthermore, from (15)

$$p^{(n)} \geq (n - s_n)\hat{x}_n$$

and so

$$p^{(n)} \geq (\bar{n} - s_{\bar{n}})\hat{x}_{\bar{n}} .$$

Now if $\bar{n} \neq n$, then a sample was taken on turn $\bar{n}+1$ so $z(\hat{x}_{\bar{n}}) < \bar{n} - s_{\bar{n}} + 1$ so that

$$p^{(n)} \geq (z(\hat{x}_{\bar{n}}) - 1)\hat{x}_{\bar{n}} . \tag{19}$$

However, if $\bar{n} = n$, then no samples have been taken up to and including turn n. This means that $z(\hat{x}) \geq n$, since with no sampling the record value is still $\hat{x}$. Thus if $z(\hat{x})$ is finite then sampling will begin after $[z(\hat{x})]$ collections, where $[y]$ = greatest integer not exceeding y, and thus we may assert (19) for sufficiently large n, provided $F(\hat{x}) < 1$, since z is finite for all such $\hat{x}$. We excluded the case $F(\hat{x}) = 1$, since then the decision procedure is pointless, and obtain for large n

$$\frac{p^{(n)} - p_n}{p^{(n)}} \leq \frac{\int_{\hat{x}}^{\hat{x}_{\bar{n}}} z(x)dx}{(z(\hat{x}_{\bar{n}}) - 1)\hat{x}_{\bar{n}}} + \frac{z(\hat{x}_{\bar{n}})}{(z(\hat{x}_{\bar{n}}) - 1)}\left[\frac{\hat{x}_n}{\hat{x}_{\bar{n}}} - 1\right] \tag{20}$$

Relation (20) holds for any continuous distribution which has a mean. We now establish sufficient conditions for the limit of the right side of (20) to be 0 for almost all $x \in X$, and thus we will have

$$p\left(\lim_{n\to\infty} \frac{p_n}{p^{(n)}} = 1\right) = 1.$$

We assume that for any ζ, $F(\zeta) < 1$, and that for any $\varepsilon > 0$,

$$\lim_{\zeta\to\infty} \frac{1 - F((1+\varepsilon)\zeta)}{1 - F(\zeta)} = 0 \ .$$

On the latter assumption, we have for all sufficiently large ζ

$$1 - F((1+\varepsilon)\zeta) \le \varepsilon(1 - F(\zeta))$$

and hence

$$1 - F((1+\varepsilon)^n\zeta) \le \varepsilon^n(1 - F(\zeta)) \ .$$

Thus

$$\frac{\int_\zeta^\infty (1 - F(x))dx}{\zeta(1 - F(\zeta))} \le \frac{\sum_{n=0}^{\infty} (1 - F((1+\varepsilon)^n\zeta))((1+\varepsilon)^{n+1} - (1+\varepsilon)^n)\zeta}{\zeta(1 - F(\zeta))}$$

$$\le \frac{\sum_{n=0}^{\infty} \varepsilon^n(1 - F(\zeta))((1+\varepsilon)^{n+1} - (1+\varepsilon)^n)\zeta}{\zeta(1 - F(\zeta))}$$

$$= \sum_{n=0}^{\infty} \varepsilon^n \ (1+\varepsilon)^{n+1} - (1+\varepsilon)^n$$

$$= \sum_{n=0}^{\infty} \varepsilon^{n+1}(1+\varepsilon)^n$$

$$= \frac{\varepsilon}{1 - \varepsilon(1+\varepsilon)} \quad \text{if } \varepsilon(1+\varepsilon) < 1 \ .$$

Thus

$$\lim_{\zeta\to\infty} \frac{\int_\zeta^\infty (1 - F(x))dx}{\zeta(1 - F(\zeta))} = 0 \qquad (21)$$

Now by L'Hôpital's rule,

$$\lim_{\zeta\to\infty} \frac{(z(\zeta) - 1)\zeta}{\int_x^\zeta z(x)dx} = \lim_{\zeta\to\infty} \frac{z(\zeta) - 1 + \zeta z'(\zeta)}{z(\zeta)} = \lim_{\zeta\to\infty} \left[1 - \frac{1}{z(\zeta)} + \frac{z'(\zeta)}{z(\zeta)}\right]$$

But $\frac{1}{z(\zeta)} \to 0$, and $\frac{z'(\zeta)}{z(\zeta)} = \frac{F(\zeta)}{\int_{-\infty}^{\zeta} F(x)dx} + \frac{(1 - F(\zeta))}{\int_\zeta^\infty (1 - F(x))dx}$.

The first term is positive, and the second term is unbounded, by (21). Thus

$$\lim_{\zeta\to\infty} \frac{(z(\zeta) - 1)\zeta}{\int_x^\zeta z(x)dx} = \infty \ .$$

However, we require this result in the equivalent form

$$\lim_{\zeta\to\infty} \frac{\int_x^\zeta z(x)dx}{(z(\zeta) - 1)\zeta} = 0 \qquad (22)$$

In order to show that with probability one $\lim_{n\to\infty} \frac{p(n) - p_n}{p(n)} = 0$ we now need only show that with probability one $\lim_{n\to\infty} \hat{x}_{\overline{n}} = \infty$ and $\lim_{n\to\infty} \frac{\hat{x}_n}{\hat{x}_{\overline{n}}} = 1$, in view of (20), (22), and the fact that z is unbounded.

We first note that since $F(\zeta) < 1$ for every ζ, then $z(\zeta)$ is finite for every ζ. Thus for any record value ζ at most $[z(\zeta)]$ collections will be made, so that sampling will eventually resume. However, once sampling has begun, it will terminate with probability one, since by (11), it terminates when a value ζ' is found such that $z(\zeta') \geq [z(\zeta)] + 1$, but in order not to find such a ζ', the largest of an arbitrarily long sequence of independent samples from x would be bounded by $z^{-1}([z(\zeta)] + 1)$ which happens with probability

$$\lim_{s\to\infty} (F(z^{-1}([z(\zeta)] + 1)))^s = 0 .$$

Thus with probability one, play proceeds by infinitely many alternate blocks of sampling and collecting. Now if we pick a sequence of numbers n_i such that for each i, n_i is in the i^{th} block of collections, then $\overline{n}_i$ will be in the $i-1^{th}$ block of collections, and since the record value is constant within each block of collections we have $\hat{x}_{\overline{n}_i} = \hat{x}_{n_{i-1}}$. Now $\hat{x}_n$ is unbounded with probability one, since with probability one infinitely many samples are taken, so $\hat{x}_{\overline{n}}$ is also unbounded.

Let us now define the random variable $\overline{\varepsilon}$ over X by

$$\overline{\varepsilon} = \sup \{\delta: \text{for infinitely many } n,\ \hat{x}_n \geq (1+\delta)\hat{x}_{\overline{n}}\}$$

Now $P(\lim_{n\to\infty} \frac{\hat{x}_n}{\hat{x}_{\overline{n}}} \neq 1) > 0$ just in case $P(\overline{\varepsilon} > 0) > 0$. We shall show that $P(\overline{\varepsilon} > 0) = 0$, and hence that $P(\lim_{n\to\infty} \frac{\hat{x}_n}{\hat{x}_{\overline{n}}} = 1) = 1$. Suppose to the contrary that $P(\overline{\varepsilon} > 0) > 0$. Then there is some $\varepsilon > 0$ such that $P(\overline{\varepsilon} > \varepsilon) > 0$. That is,

$$P(\sup\{\delta: \text{for infinitely many } n,\ \hat{x}_n \geq (1+\delta)\hat{\underline{x}}_n\} > \varepsilon)$$

$$= P(\text{for infinitely many } n,\ \hat{x}_n \geq (1+\ )\hat{\underline{x}}_n)$$

$$> 0$$

Now with probability 1, $\hat{x}_n$ is unbounded, so we may exclude those $\vec{x} \in X$ for which $\hat{x}_n$ is bounded, and choose values n_i as before. Then

$$P(\text{for infinitely many } i\ \hat{x}_{n_{i+1}} > (1+\varepsilon)\hat{x}_{n_i}) > 0$$

But we shall now show that

$$\lim_{\zeta\to\infty} P(\hat{x}_{n_{i+1}} > (1+\varepsilon)\zeta \,|\, \hat{x}_{n_i} = \zeta) = 0$$

and hence have our desired contradiction. The conditional density function of $\hat{x}_{n_{i+1}}$ given $\hat{x}_{n_i} = \zeta$, is

$$\frac{f(\hat{x}_{n_{i+1}})}{1 - F(z^{-1}([z(\zeta)] + 1))} \quad \text{for } \hat{x}_{n_{i+1}} > z^{-1}([z(\zeta)] + 1)$$

Thus

$$P(\hat{x}_{n_{i+1}} > (1+\varepsilon)\zeta \,|\, \hat{x}_{n_i} = \zeta) = \frac{1 - F(\max((1+\varepsilon)\zeta, z^{-1}([z(\zeta)] + 1)))}{1 - F(z^{-1}([z(\zeta)] + 1))}$$

We note that

$$z(\zeta+\varepsilon) = \frac{\int_{\infty}^{\zeta} {}_{\varepsilon}F(y)dy}{\int_{\zeta}^{\infty} {}_{\varepsilon}(1 - F(y))dy} \geq \frac{\int_{-\infty}^{\zeta} F(y)dy + \int_{\zeta}^{\zeta} {}_{\varepsilon}F(y)dy}{\int^{\infty}(1 - F(y))dy} > z(\zeta) + \frac{F(\zeta)}{\int_{\zeta}^{\infty}(1 - F(y))dy}$$

Thus if ζ is sufficiently large that

$\frac{\int_{\zeta}^{\infty}(1 - F(y))dy}{F(\zeta)} < \varepsilon$, then $z(\zeta+\varepsilon) > z(\zeta) + 1$ and

$\zeta+\varepsilon = z^{-1}(z(\zeta+\varepsilon)) > z^{-1}(z(\zeta)+1) \geq z^{-1}([z(\zeta)]+1)$. Thus for large ζ,

$$\frac{1 - F(\max((1+\varepsilon)\zeta, z^{-1}([z(\zeta)]+1))}{1 - F(z^{-1}([z(\zeta)]+1))} < \frac{1 - F((1+\varepsilon)\zeta)}{1 - F(\zeta+\varepsilon)} < \frac{1 - F((1+\delta)(\zeta+\varepsilon))}{1 - F(\zeta+\varepsilon)}$$

for suitable δ and large ζ. But the right term approaches 0, by hypothesis.

This completes the proof that $\lim_{n\to\infty} \frac{p^{(n)} - p_n}{p^{(n)}} = 0$ with probability one,

for suitable distributions.

We note that if the random variable is bounded, we also have the above result, with no assumptions other than that of a continuous distribution, and the existence of a value x_{max} such that $F(x_{max}) = 1$ and $F(x) < 1$ for $x < x_{max}$. To show this, we note that

$$\lim_{\zeta \to x_{max}} \frac{\int_{\zeta}^{x_{max}} (1 - F(x))dx}{\zeta(1 - F(\zeta))} = 0$$

and that with probability one $\lim_{n\to\infty} x_{\overline{n}} = x_{max}$ and $\lim_{n\to\infty} \frac{x_n}{x_{\overline{n}}} = 1$, so that we obtain the result directly from (20).

In summary, we have shown that $\lim_{n\to\infty} \frac{p_n}{p^{(n)}} = 1$ for any bounded random variable x with a continuous density function, or for any random variable with mean and continuous density function satisfying

(1) $F(\zeta) < 1$ for all δ.

(2) $\lim_{\zeta\to\infty} \frac{1 - F((1+\varepsilon)\zeta)}{1 - F(\zeta)} = 0$ for all $\varepsilon > 0$.

We note that many standard distributions such as the normal and exponential distributions have the properties (1) and (2).

DISTRIBUTION LIST

(One copy unless otherwise noted)

Technical Library
Director Defense Res. & Eng.
Room 3C-128, The Pentagon
Washington, D.C. 20301

Defense Documentation Center 20
Cameron Station
Alexandria, Virginia 22314

Chief of Naval Research 2
Department of the Navy
Washington 25, D.C.
Attn: Code 437, Information
Systems Branch

Director, Naval Research Laboratory 6
Technical Information Officer
Washington 25, D.C.
Attention: Code 2000

Commanding Officer 10
Office of Naval Research Branch Office
Box 39, Fleet Post Office
New York, New York 09510

Commanding Officer
ONR Branch Office
207 West 24th Street
New York 11, New York

Office of Naval Research Branch
Office
495 Summer Street
Boston, Massachusetts 02110

Naval Ordnance Laboratory
White Oaks, Silver Spring 19
Maryland
Attn: Technical Library

David Taylor Model Basin
Washington,D.C. 20007
Attn: Code 042, Technical Library

Naval Electronics Laboratory
San Diego 52, California
Attn: Technical Library

Dr. Daniel Alpert, Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois

Air Force Cambridge Research Labs
Laurence C. Hanscom Field
Bedford, Massachusetts
Attn: Research Library, CRMXL R

U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448
Attn: G. H. Gleissner, Code K4
Asst. Dir. for Computation

National Bureau of Standards
Data Processing Systems Division
Room 239, Building 10
Washington 25, D.C.
Attn: A. K. Smilow

George C. Francis
Computing Laboratory, BRL
Aberdeen Proving Ground, Maryland

Office of Naval Research
Branch Office, Chicago
230 North Michigan Avenue
Chicago, Illinois 60601

Commanding Officer
ONR Branch Office
1030 E. Green Street
Pasadena, California

Commanding Officer
ONR Branch Office
1076 Mission Street
San Francisco, California 94103

DISTRIBUTION LIST (Concluded)

The University of Michigan
Department of Philosophy
Attn: Professor A. W. Burks

National Physical Laboratory
Teddington, Middlesex, England
Attn: Dr. A. M. Uttley, Supt.
Autonomics Division

Commanding Officer
Harry Diamond Laboratories
Washington, D.C. 20438
Attn: Library

Commanding Officer and Director
U. S. Naval Training Device Center
Orlando, Florida 32813
Attn: Technical Library

Department of the Army
Office of the Chief of Research
and Development
Pentagon, Room 3D442
Washington 25, D.C.
Attn: Mr. L. H. Geiger

National Security Agency
Fort George G. Meade, Maryland
Attn: Librarian, C-332

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington 73, Massachusetts
Attn: Library

Office of Naval Research
Washington 25, D.C.
Attn: Code 432

Dr. Kenneth Krohn
Krohn Rhodes Research Institute, Inc.
328 Pennsylvania Avenue, S. E.
Washington 13, D. C.

Dr. Larry Fogel
Decision Science, Inc.
6508 Pacific Highway
San Diego, California

National Bureau of Standards
Applications Engineering Section
Washington 25, D. C.
Attn: Miss Mary E. Stevens

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Logic of Computers Group The University of Michigan Ann Arbor, Michigan 48104	Unclassified
	2b. GROUP

3. REPORT TITLE

A CLASS OF SEQUENTIAL SAMPLING PROBLEMS ARISING IN CERTAIN LEARNING SITUATIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (Last name, first name, initial)

Bainbridge, Edwin S.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1967	18	-

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Nonr 1224(21)	03105-49-T
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC.
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Department of the Navy Washington 25, D.C.

13. ABSTRACT

A strategist is to decide on each of n turns whether to take a sample from a certain fixed random variable, and receive the outcome as payoff, or to receive as payoff the largest value of the random variable he has discovered so far. The expected total payoff for the n turns is to be maximized. It is shown that the following decision procedure is the solution.

Sample until $\frac{\int_{-\infty}^{\hat{x}} F(x)dx}{\int_{\hat{x}}^{\infty}(1-F(x))dx}$ exceeds the number of turns remaining where $\hat{x}$ is the current record value, and $F(x)$ is the distribution function of the random variable.

A strategy for an indefinite number of turns is described, and for suitable distributions it is shown that the limit of the ratio of the payoff accumulated by this strategy in n turns to the payoff accumulated by the optimal strategy for n turns is one, with probability one.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
statistics						
sequential sampling						
dynamic programming						
trial and error learning						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

Unclassified

Security Classification