*The Annals of Statistics*
1976, Vol. 4, No. 4, 807-816

# SEQUENTIAL SEARCH WITH RANDOM OVERLOOK PROBABILITIES[1]

BY GAINEFORD J. HALL, JR.

*The University of Texas at Austin*

Suppose an object is hidden in one of $N$ boxes. The initial prior probability that it is hidden in box $i$ is known to the searcher, who once a day must choose a box to be searched. The probability that the $j$th search of box $i$ will be unsuccessful, given that the object is in box $i$, is described by a random variable $\alpha_{ij}$. Assuming that each search of box $i$ costs $c_i > 0$, sufficient conditions on the joint distribution of the $\{\alpha_{ij}\}_{j=1}^{\infty}$ are found in order to guarantee that a particular search rule (analogous to earlier search rules found independently by R. Bellman, D. Blackwell, W. Black and J. Kadane) is optimal with respect to minimizing the total expected search cost of finding the object. An extension of a search problem studied by C. Sweat is also treated.

**1. Introduction and summary.** An object is hidden in one of $N$ boxes, labeled $1, \cdots, N$, where it remains until found. It is in box $i$ with prior probability $p_i^1 = 0$, where $\sum_{i=1}^N p_i^1 = 1$. Knowing these probabilities, a searcher selects a box to be searched each day. The conditional probability that the $j$th search of box $i$ discovers the objects, given that it is in box $i$, is a random variable $\alpha_{ij}$, called a *random overlook probability* or *overlook random variable*. For each $i$, the random variables $\alpha_{i1}, \alpha_{i2}, \cdots$ have a joint distribution which the searcher also knows, while the sequences $\{\alpha_{1j}\}_{j=1}^{\infty}, \cdots, \{\alpha_{Nj}\}_{j=1}^{\infty}$ are independent.

Two forms of the problem are considered. In the first, each search of box $i$ costs $c_i > 0$, $1 \leqq i \leqq N$. In the second, there is no search cost but there is a positive probability $1 - \beta_i$ that a search of box $i$ will result in termination of the search process without finding the object. The searcher receives a reward of one unit if he finds the object before termination, and zero reward otherwise. Thus $\beta_i$, $0 < \beta_i < 1$, acts as a discount factor.

Consider the problem with costs. Suppose that the searcher has already searched unsuccessfully for the object for $n - 1$ days and assume that box $l$ has been searched $m(l) \geqq 0$ times, so that $\sum_{l=1}^N m(l) = n - 1$, $1 \leqq l \leqq N$. The searcher is aware of the values $t_{l1}, \cdots, t_{l,m(l)}$ which the random overlook probabilities $\alpha_{l1}, \cdots, \alpha_{l,m(l)}$, respectively, had assumed for each box $l$. If the searcher selects box $i$ to be searched next, he pays cost $c_i$ and learns the actual value of $\alpha_{i,m(i)+1}$ during this search of box $i$.

---

Received January 1974; revised December 1975.

[1] This paper is based on part of the author's Ph. D. dissertation. The author gratefully acknowledges the support of the National Science Foundation while a graduate student at UCLA, under NSF Traineeship GZ1547.

*AMS* 1970 *subject classification*. Primary 90B40.

*Key words and phrases*. Sequential search, random overlook probability, optimal policy, dynamic programming, decision model.

If the searcher finds the object, search is terminated and no more cost is incurred. Otherwise the searcher uses the observed value $t = \alpha_{i,m(i)+1}$ to revise his opinion of the object's location (via Bayes' rule) and search continues. The value $t$ is the probability that the searcher will fail to find the object on the $(m(i) + 1)^{st}$ search of box $i$, given it is in box $i$. The searcher must find an optimal policy to minimize the total expected search cost of finding the object. If for each $i$ $\alpha_{ij} \equiv \alpha_i$ is a constant independent of $j$ then this search problem reduces to the one solved by Blackwell [3] and Black [2]. If for each $i$ $\{\alpha_{ij}\}_{j=1}^{\infty}$ is a sequence of (possibly different) constants then this search model reduces to the model considered by Kadane [8]. We could treat the search problem wherein the cost for the $j$th search of box $i$, $c_{ij}$, depends on $j$, as in Kadane, but the proof for the more general cost structure of Kadane is similar to the proof given.

In the discounted reward problem, if a search of box $i$ is unsuccessful, the search process continues with probability $\beta_i$ and the searcher's opinion of the object's location is revised via Bayes' rule. The searcher must find an optimal policy to maximize the total expected discounted reward, or equivalently to maximize the probability of eventually finding the object. If for each $i$ $\alpha_{ij} \equiv \alpha_i$ is a constant independent of $j$ then this problem reduces to that of Sweat [11].

We obtain the following results in this paper: For the search problem with cost we find necessary and sufficient conditions on $\{\alpha_{ij}\}_{j=1}^{\infty}$, $1 \leqq i \leqq N$, such that the minimal total expected search cost is finite. Under a certain monotonicity condition on the joint distribution of $\{\alpha_{ij}\}_{j=1}^{\infty}$, $1 \leqq i \leqq N$, we show that an optimal search rule exists which agrees with the optimal rules found by Blackwell, Black and Kadane when the $\alpha_{ij}$ are constant. (If $\{\alpha_{ij}\}_{j=1}^{\infty}$ is a sequence of constants, our monotonicity condition reduces to that of Kadane for the case $c_{ij} \equiv c_i$, independent of $j$.) We give an example in which the monotonicity condition fails to hold and this analog of the Blackwell–Black–Kadane rule is not optimal. For the search problem with discounted reward, we show that under the same monotonicity condition there is an optimal search rule which is the appropriate analog of the Sweat search rule. Lastly, we shall give some examples illustrating the theory.

## 2. The dynamic programming model for the search problem with cost.

In order to show the existence of an optimal search policy, we formulate the search problem with cost as a negative dynamic programming problem (cf. Strauch [10]). To do this we must define the state space $S$, the action space $A$, the transition probability $q(\cdot \mid \cdot, \cdot): S \times A \to S$ and the cost function $c: S \times A \times S \to R$.

The action space is $A = \{1, 2, \cdots, N\}$. To define $S$, first let $\Delta^N = \{\mathbf{p} = (p_1, \cdots, p_N): p_i \geqq 0,\ 1 \leqq i \leqq N, \text{ and } \sum_{i=1}^{N} p_i = 1\}$. Thus $\Delta^N$ is the set of all location vectors $\mathbf{p}$ for the hidden object. Define $s^*$, an isolated point, to be the state into which the system is mapped once the object has been found. For $i \in A$ and $t \in [0, 1]$, define the map $T(i, t): \Delta^N \to \Delta^N \cup \{s^*\}$ as follows. If $\mathbf{p} \in \Delta^N$, $T(i, t)\mathbf{p} = \mathbf{p}'$ where $p_i' = tp_i/(1 - (1 - t)p_i)$ and $p_l' = p_l/(1 - (1 - t)p_i)$ for

$l \neq i$ (here we tacitly assume $1 - (1 - t)p_i \neq 0$). If $\mathbf{p}$ is the prior location vector at some stage, action $i$ is taken, the value $t$ is observed and the search of box $i$ is unsuccessful, then $T(i, t)\mathbf{p}$ is the posterior location vector for the object, computed by Bayes' rule. If $t = 0$ and $\mathbf{p} = \mathbf{e}^i$ ($\mathbf{e}^i$ is the vertex of $\Delta^N$ with 1 in the $i$th coordinate and 0 elsewhere) define $T(i, 0)\mathbf{e}^i = s^*$, since if the overlook is zero during a search of box $i$ when $\mathbf{p} = \mathbf{e}^i$, the object is sure to be found.

Now define $S_1 = \Delta^N$ and for $n \geqq 2$, define $S_n = \{s_n = (\mathbf{p}^1, i_1, \mathbf{p}^2, i_2, \cdots, i_{n-1}, \mathbf{p}^n)$: $\forall\, m$, $1 \leqq m \leqq n-1$, $\mathbf{p}^{m+1} \in \Delta^N$ and $\exists\, u_m \in [0, 1]$ such that $\mathbf{p}^{m+1} = T(i_m, u_m)\mathbf{p}^m\}$. Finally, $S = \bigcup_{n=1}^{\infty} S_n \cup \{s^*\}$, the disjoint union with the usual topology and Borel sets. For notation, if $i_m = l \in A$, let $t_{l,\nu_m(l)} = u_m$, where $\nu_m(l) = \nu_m(l; i_1, \cdots, i_{m-1})$ is the number of times that $l$ occurs in $i_1, \cdots, i_m$, $1 \leqq m \leqq n-1$.

Let $\mu_{i1}$ denote the (marginal) distribution of $\alpha_{i1}$ and for each $m \geqq 1$ let $\mu_{i,m+1}$ be a transition probability of $[0, 1]^m$ into $[0, 1]$ which is a conditional distribution of $\alpha_{i,m+1}$ given $\alpha_{i1}, \cdots, \alpha_{im}$, $1 \leqq i \leqq N$. To define the transition probability $q$, let $s \in S$ and $i \in A$. If $s = s_n \in S_n$ for some $n$, let $t_{i1}, \cdots, t_{i,m(i)}$ denote the observed values of $\alpha_{i1}, \cdots, \alpha_{i,m(i)}$ occurring in $s_n$, as before, where $m(i) = m(i; s_n)$ is the number of times $i$ is searched in $s_n$. For convenience of notation write $\mu_{i,m(i)+1}(\cdot \mid s_n) \equiv \mu_{i,m(i)+1}(\cdot \mid t_{i1}, \cdots, t_{i,m(i)})$. Then define $q(s^* \mid s_n, i) = p_i^n[1 - E(\alpha_{i,m(i)+1} \mid s_n)]$ where $E(\alpha_{i,m(i)+1} \mid s_n) = \int_0^1 t\mu_{i,m(i)+1}(dt \mid s_n)$. For $B$ a Borel subset of $\Delta^N$, define $q(\{(s_n, i)\} \times B \mid s_n, i) = \int_C [1 - (1 - t)p_i^n]\mu_{i,m(i)+1}(dt \mid s_n)$, where $C = \{t \in [0, 1]: T(i, t)\mathbf{p}^n \in B\}$. Thus $q(S_{n+1} \mid s_n, i) = 1 - q(s^* \mid s_n, i)$. Lastly, define $q(s^* \mid s^*, i) = 1$, for all $i$. Thus for $s_n \in S_n$, $q(s^* \mid s_n, i)$ is the probability that the object is found during the $(m(i) + 1)^{st}$ search of box $i$ and $q(\{(s_n, i)\} \times B \mid s_n, i)$ is the probability that the object is not found and the posterior location vector will lie in $B$. To define the cost function $c$, let $c(s_n, i, s') = c_i$, for $s_n \in S_n$, $s' \in S$, $i \in A$, and let $c(s^*, i, s') = 0$.

Thus we have completely specified the decision model for our problem. By results of Strauch [10], there exists a stationary optimal policy. Here, a policy $\sigma = (\sigma_n)_{n=1}^{\infty}$ is a sequence of maps $\sigma_n: S_n \to A$ where $\sigma_n(s_n)$ is the action taken at time $n$ at state $s_n$. Furthermore, $\varphi_\sigma(\mathbf{p})$ denotes the expected search cost under policy $\sigma$ at state $\mathbf{p}$ and $\varphi \equiv \inf_\sigma \varphi_\sigma$ is the total minimal expected search cost. The first theorem gives necessary and sufficient conditions on $\{\alpha_{ij}\}_{j=1}^{\infty}$ in order that $\varphi(\mathbf{p}) < \infty$.

THEOREM 1. *Suppose that for the initial state* $\mathbf{p}$ *we have* $p_i > 0$, *for each* $i$. *Then* $\varphi(\mathbf{p}) < \infty$ *if and only if* $\sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij}) < \infty$, $1 \leqq i \leqq N$.

PROOF. First we show that if $\sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij}) < \infty$ for each $i$ then $\varphi(\mathbf{p}) < \infty$ for all $\mathbf{p} \in \Delta^N$. Let $\pi$ be the policy which searches boxes $1, 2, \cdots, N$ cyclically until the object is found. If the object is in box $i$, so that $\mathbf{p} = \mathbf{e}^i$, then the expected number of searches required to find the object (by always searching in box $i$) is $1 + \sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij})$. Hence $\varphi(\mathbf{e}^i) \leqq \varphi_\pi(\mathbf{e}^i) \leqq (\sum_{l=1}^{N} c_l)(1 + \sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij}))$. Thus if $\mathbf{p} \in \Delta^N$ is the initial prior location vector, then $\varphi(\mathbf{p}) \leqq \varphi_\pi(\mathbf{p}) \leqq (\sum_{l=1}^{N} c_l)(\sum_{i=1}^{N} p_i[1 + \sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij})])$. Hence if

$$\max_{1\leqq i\leqq N} \sum_{n=1}^{\infty} E(\prod_{j=1}^{n} \alpha_{ij}) < \infty$$

then $\varphi(\mathbf{p}) < \infty$ for all $\mathbf{p} \in \Delta^N$.

Conversely, suppose that $p_i > 0$ for $1 \leqq i \leqq N$ and that $\varphi(\mathbf{p}) < \infty$. Now $\varphi$ is concave on $\Delta^N$, since $\varphi(\mathbf{p}) = \min_\sigma \varphi_\sigma(\mathbf{p}) = \min_\sigma \sum_{i=1}^N p_i \varphi_\sigma(\mathbf{e}^i) \geqq \sum_{i=1}^N p_i \min_\sigma \varphi_\sigma(\mathbf{e}^i) = \sum_{i=1}^N p_i \varphi(\mathbf{e}^i)$ for all $\mathbf{p} \in \Delta^N$. Therefore if each $p_i > 0$, $1 \leqq i \leqq N$, we must have $\sum_{n=1}^\infty E(\prod_{j=1}^n \alpha_{ij}) < \infty$, for each $i$. □

From now on we shall assume that $\sum_{n=1}^\infty E(\prod_{j=1}^n \alpha_{ij}) < \infty$ for each $i$. We also assume that $p_i > 0$, $1 \leqq i \leqq N$, for the initial location vector $\mathbf{p}$.

**3. The solution to the search problem with cost.** In this section we find conditions on the joint distribution of $\{\alpha_{ij}\}_{j=1}^\infty$, $1 \leqq i \leqq N$, under which the appropriate analog of the Blackwell–Black–Kadane search policy is optimal. We first prove the following theorem.

THEOREM 1. (i) *In the search problem with cost, suppose that the following inequality is satisfied for all $i$, $2 \leqq i \leqq N$, and all $m \geqq 1$:*

$$(1) \qquad p_1[1 - E\alpha_{11}]/c_1 \geqq p_i(\prod_{j=1}^{m-1} \alpha_{ij})[1 - E(\alpha_{im} \mid \alpha_{i1}, \cdots, \alpha_{i,m-1})]/c_i\,, \quad \text{a.s.}$$

*Then if $\sigma$ is any search policy with $\sigma_1(\mathbf{p}) \neq 1$, there is a policy $\sigma'$ with $\sigma_1'(\mathbf{p}) = 1$ and $\varphi_\sigma(\mathbf{p}) \geqq \varphi_{\sigma'}(\mathbf{p})$.* (ii) *If $\mathbf{p}$ is such that $p_1[1 - E\alpha_{11}]/c_1 > p_i[1 - E\alpha_{i1}]/c_i$, $2 \leqq i \leqq N$, and* (1) *holds, then $\varphi_\sigma(\mathbf{p}) > \varphi_{\sigma'}(\mathbf{p})$, where $\sigma$, $\sigma'$ are as in* (i).

PROOF. We show that for any search rule $\sigma$ such that $\varphi_\sigma(\mathbf{p}) < \infty$ and $\sigma_1(\mathbf{p}) = h \neq 1$ there is a rule $\sigma'$ such that $\sigma_1'(\mathbf{p}) = 1$ and $\varphi_\sigma(\mathbf{p}) - \varphi_{\sigma'}(\mathbf{p}) \geqq c_h p_1[1 - E\alpha_{11}] - c_1 p_h[1 - E\alpha_{h1}]$, from which both (i) and (ii) follow. Intuitively, (1) states that box 1 is the "most attractive" box to search at state $\mathbf{p}$, and if some other box is searched, box 1 remains the most attractive.

Now let $\{t_{ij}\}_{j=1}^\infty$ be a sample path of $\{\alpha_{ij}\}_{j=1}^\infty$, $1 \leqq i \leqq N$, and let the sequence of decisions which $\sigma$ prescribes for this sample path be denoted by $i_1, i_2, \cdots, i_{m-1}, i_m, i_{m+1}, \cdots$ where by assumption $i_1 = h \neq 1$. Suppose that $i_m = 1$ and that $i_k \neq 1$ for $1 \leqq k \leqq m - 1$. Let $\sigma'$ be the policy which prescribes for this sample path of overlook probabilities the actions $1, i_1, \cdots, i_{m-1}, i_{m+1}, i_{m+2}, \cdots$. Notice that this determines a well-defined policy $\sigma'$. We find the difference of the conditional expected values of the search cost under policies $\sigma$ and $\sigma'$, given $\alpha_{ij} = t_{ij}$, $j \geqq 1$, $1 \leqq i \leqq N$.

Let $1 \leqq k \leqq m - 1$, and notice that the difference of the conditional expectations of total cost for the sequences of actions $i_1, i_2, \cdots, i_{k-1}, i_k, 1, i_{k+2}, \cdots$ and $i_1, i_2, \cdots, i_{k-1}, 1, i_k, i_{k+2}, \cdots$ equals the difference in conditional expected cost at times $k$ and $k + 1$ since before time $k$ and after time $k + 1$ the conditional expected costs are the same. Thus, the conditional difference given that the object is not found before time $k$ is

$$(2) \qquad \begin{aligned} c_{i_k} + [1 - p_{i_k}^k(1 - t_{i_k,\nu_k(i_k)})]c_1 - (c_1 + [1 - p_1^k(1 - t_{11})]c_{i_k}) \\ = p_1^k(1 - t_{11})c_{i_k} - p_{i_k}^k(1 - t_{i_k,\nu_k(i_k)})c_1\,, \end{aligned}$$

where $\nu_k(l)$ is the number of times $l$ occurs among $i_1, \cdots, i_k$. By the definition of the map $T(i, t)$,

$$p_i^k = p_i(\prod_{j=1}^{\nu_{k-1}(i)} t_{ij})[\sum_{l=1}^N p_l(\prod_{j=1}^{\nu_{k-1}(l)} t_{lj})]^{-1}, \qquad 1 \leqq i \leqq N. \tag{3}$$

The probability that the object is not found at stages $1, 2, \cdots, k-1$ is, given $\alpha_{ij} = t_{ij}, j \geqq 1, 1 \leqq i \leqq N$,

$$\sum_{l=1}^N p_l \prod_{j=1}^{\nu_{k-1}(l)} t_{lj}. \tag{4}$$

The difference of conditional expected costs under $\sigma$ and $\sigma'$ is found by using the sum of differences (2) by moving action 1 each step one place further toward stage 1. Combining this with (3) and (4) we get for the conditional expected difference

$$\sum_{k=1}^{m-1} [p_1(1 - t_{11})c_{i_k} - p_{i_k}(\prod_{j=1}^{\nu_{k-1}(i_k)} t_{i_k,j})(1 - t_{i_k,\nu_k(i_k)})c_1]. \tag{5}$$

Let $J_k$ be the random variable which is obtained from the $k$th term of (5) by substitution of $\{\alpha_{ij}\}$ for $\{t_{ij}\}$ and by substitution of the action which $\sigma$ prescribes at time $k$ for $i_k$. Let $\xi$ be the first time $\sigma$ prescribes action 1. Since $p_1 > 0$ and $\varphi_\sigma(\mathbf{p}) < \infty$, $2 \leqq \xi < \infty$ a.s., under policy $\sigma$. Thus

$$\begin{aligned}\varphi_\sigma(\mathbf{p}) - \varphi_{\sigma'}(\mathbf{p}) &= E_\sigma[\sum_{m=2}^\infty I_{\{\xi=m\}} \sum_{k-1}^{m-1} J_k] \\ &= E_\sigma[\sum_{k=1}^\infty I_{\{\xi>k\}} J_k] \geqq E_\sigma[J_1],\end{aligned}$$

by (1) and the independence of the $N$ processes, where $I_B$ is the indicator function of event $B$. Note that $I_{\{\xi<k\}}$ is a function only of $\alpha_{ij}$ for $1 \leqq j \leqq \nu_{k-1}(i)$, $1 \leqq i \leqq N$. Since $E_\sigma[J_1] = p_1(1 - E\alpha_{11})c_h - p_h(1 - E\alpha_{h1})c_1$, the conclusion of the theorem follows. □

Next we define the "weak monotonicity condition."

DEFINITION 1. Let $\alpha_1, \alpha_2, \alpha_3, \cdots$ be a sequence of $[0, 1]$-valued random variables. We say that $\{\alpha_j\}_{j=1}^\infty$ satisfies the weak monotonicity condition (or W.M.C.) if for all $m \geqq 1$,

$$1 - E\alpha_1 \geqq (\prod_{j=1}^{m-1} \alpha_j)[1 - E(\alpha_m \mid \alpha_1, \cdots, \alpha_{m-1})], \quad \text{a.s.} \tag{6}$$

The W.M.C. is a kind of "local optimality" condition, as shown by the next theorem.

THEOREM 2. *Suppose that in the search problem with cost for each box $i$, $1 \leqq i \leqq N$, the overlook random variables $\{\alpha_{ij}\}_{j=1}^\infty$ satisfy the* W.M.C. (6). *Let box $i$ be such that $p_i[1 - E\alpha_{i1}]/c_i = \max_{1\leqq l\leqq N} p_l[1 - E\alpha_{l1}]/c_l$. Then if $\sigma$ is any policy with $\sigma_1(\mathbf{p}) \neq i$, there is a policy $\sigma'$ with $\sigma_1'(\mathbf{p}) = i$ and $\varphi_\sigma(\mathbf{p}) \geqq \varphi_{\sigma'}(\mathbf{p})$.*

PROOF. The conclusion follows immediately from (i) of Theorem 1. □

The monotonicity condition referred to in the introduction is the following.

DEFINITION 2. The sequence $\{\alpha_j\}_{j=1}^\infty$ of $[0, 1]$-valued random variables satisfies the strong monotonicity condition (or S.M.C.) if $(\prod_{j=1}^{m-1} \alpha_j)[1 - E(\alpha_m \mid \alpha_1, \cdots, \alpha_{m-1})]$ is monotone nonincreasing in $m$, a.s.

Notice that the S.M.C. essentially implies that the W.M.C. is satisfied at each stage and that if $\{\alpha_j\}_{j=1}^{\infty}$ are i.i.d. [0, 1]-valued random variables the S.M.C. is automatically satisfied. The search policy we are interested in is the following.

DEFINITION 3. Define $f = (f_n)_{n=1}^{\infty}$ as follows. If $s = s_n \in S_n$, let $f_n(s_n)$ be any action $i$ achieving

$$(7) \qquad p_i^n[1 - E(\alpha_{i,m(i)+1} \mid s_n)]/c_i = \max_{1 \leqq l \leqq N} p_l^n[1 - E(\alpha_{l,m(l)+1} \mid s_n)]/c_l \,,$$

where $\mathbf{p}^n$ is the last coordinate of $s_n$ and $m(l)$ is the number of times box $l$ has been searched in $s_n$. If $s = s^*$, $f_n(s^*)$ is arbitrary. This is the Blackwell–Black–Kadane (or B.B.K.) policy.

Roughly speaking, this policy always selects that box which has maximum present probability of success per unit cost of search, i.e., it always selects the "most attractive" box.

THEOREM 3. *Suppose that in the search problem with cost, for each i,* $1 \leqq i \leqq N$, *the overlook random variables* $\{\alpha_{ij}\}_{j=1}^{\infty}$ *satisfy the* S.M.C. *Then the* B.B.K. *policy is optimal,* i.e., $\varphi \equiv \varphi_f$ *on* $\Delta^N$.

PROOF. Note that on the event $\{\prod_{j=1}^{k} \alpha_{ij} > 0\}$, $\{\alpha_{ij}\}_{j=k+1}^{\infty}$ satisfies the S.M.C., for any $k \geqq 1$. Hence we may use the fact that an optimal policy exists together with the proof of Theorem 1 at each stage $n$ to conclude that the B.B.K. rule is optimal. □

**4. The search problem with discounted reward.** In this section we study the extension of the Sweat problem described in the introduction. Note that the problem can be extended to that in which the discount factor $\beta_{ij}$ depends on the $j$th search of box $i$, but again, the solution is similar to that given.

To formulate the discounted reward search problem as a dynamic programming problem, define the state space $S$, action space $A$ and transition probability $q(\cdot \mid \cdot, \cdot)$ as in Section 2. To define the reward $r: S \times A \times S \to \mathbb{R}$, let $r(s^*, i, s') \equiv 0$ for all $s' \in S$ and $r(s_n, i, s') = (\prod_{\nu=1}^{n-1} \beta_{i_\nu}) p_i^n[1 - E(\alpha_{i,m(i)+1} \mid s_n)]$, as in our earlier notation, where $i_1, \cdots, i_{n-1}$ is the sequence of searches in $s_n \in S_n$. Then we know from results of Blackwell [4] or alternatively from Hinderer [7] that an optimal policy with respect to the total maximal expected discounted reward exists. Let $I_\sigma(\mathbf{p})$ denote the expected discounted reward under policy $\sigma$ and $I(\mathbf{p}) = \sup_\sigma I_\sigma(\mathbf{p})$ be the maximal expected discounted reward.

THEOREM 1. (i) *In the search problem with discounted reward, suppose that the following inequality is met for all i,* $2 \leqq i \leqq N$, *and all* $m \geqq 1$:

$$(1) \qquad p_1[1 - E\alpha_{11}]/(1 - \beta_1) \geqq p_i(\textstyle\prod_{j=1}^{m-1} \alpha_{ij})[1 - E(\alpha_{im} \mid \alpha_{i1}, \cdots, \alpha_{i,m-1})]/(1 - \beta_i) \,, \quad \text{a.s.}$$

*Then if* $\sigma$ *is any policy with* $\sigma_1(\mathbf{p}) \neq 1$, *there is a policy* $\sigma'$ *with* $\sigma_1'(\mathbf{p}) = 1$ *and* $I_\sigma(\mathbf{p}) \leqq I_{\sigma'}(\mathbf{p})$. (ii) *If* $\mathbf{p}$ *is such that* $p_1[1 - E\alpha_{11}]/(1 - \beta_1) > p_i[1 - E\alpha_{i1}]/(1 - \beta_i)$, $2 \leqq i \leqq N$, *and* (1) *holds, then* $I_\sigma(\mathbf{p}) < I_{\sigma'}(\mathbf{p})$, *where* $\sigma$, $\sigma'$ *are as in* (i).

PROOF. Analogously to the proof of Theorem 1, Section 3, we show that for any search rule $\sigma$ such that $\sigma_1(\mathbf{p}) = h \neq 1$ there is a rule $\sigma'$ such that $\sigma_1'(\mathbf{p}) = 1$ and $I_{\sigma'}(\mathbf{p}) - I_\sigma(\mathbf{p}) \geqq p_1[1 - E\alpha_{11}](1 - \beta_h) - p_h[1 - E\alpha_{h1}](1 - \beta_1)$, from which both (i) and (ii) follow. We indicate the appropriate changes in the proof of the earlier Theorem 1.

Given $\{\alpha_{ij}\} = \{t_{ij}\}$, it is easy to show that the probability of finding the object under the sequence $i_1, \cdots, i_{k-1}, 1, i_k, i_{k+2}, \cdots$ minus the probability under the sequence $i_1, \cdots, i_{k-1}, i_k, 1, i_{k+2}, \cdots$, given that the object is not found before time $k$ and the search does not terminate before $k$ is

$$(2) \qquad p_1^k(1 - t_{11})\{1 - \beta_{i_k}\} - p_{i_k}^k(1 - t_{i_k, \nu_k(i_k)})\{1 - \beta_1\} .$$

This is the expression in the right-hand side of (2), Section 3, with $1 - \beta_l$ for $c_l$. The probability that the object is not found and the search does not terminate before stage $k$ is

$$(3) \qquad (\prod_{\nu=1}^{k-1} \beta_{i_\nu}) \sum_{l=1}^N p_l^1 \prod_{j=1}^{\nu_{k-1}(l)} t_{lj} .$$

The posterior probabilities are in (3), Section 1. The remainder of the proof is the same. □

We shall leave to the reader the task of restating Theorem 2, Section 3 for the Sweat search problem, where $c_l$ is replaced by $1 - \beta_l$ and $\varphi_\sigma$ is replaced by $I_\sigma$. The policy of interest for the Sweat problem is the following.

DEFINITION 1. The Sweat search policy $\hat{\sigma} = (\hat{\sigma}_n)_{n=1}^\infty$ is defined by setting $\hat{\sigma}_n(s_n)$ equal to any action $i$ such that

$$(4) \qquad p_i^n[1 - E(\alpha_{i,m(i)+1} \mid s_n)]/(1 - \beta_i) = \max_{1 \leqq l \leqq N} p_l^n[1 - E(\alpha_{l,m(l)+1} \mid s_n)]/(1 - \beta_l) ,$$

where $s_n \in S_n$, $n \geqq 1$.

THEOREM 3. *Suppose that in the search problem with discounted reward for each $i$ the overlook random variables $\{\alpha_{ij}\}_{j=1}^\infty$ satisfy the* S.M.C. *of Section* 2. *Then the Sweat search rule is optimal*, i.e., $I \equiv I_{\hat{\sigma}}$ *on* $\Delta^N$.

PROOF. Follows immediately from Theorem 1 of this section and the fact that there is an optimal policy. □

**5. Applications and examples.** In this section we give several interesting applications and examples of the preceding theorems.

(a) *Searching for a gold coin.* Given are $N$ boxes, each box $i$ containing a known number $n_i > 0$ of brass coins. A box is chosen at random, box $i$ with probability $p_i^1 > 0$, and one brass coin is replaced with a gold one. A searcher, who knows the initial vectors $\mathbf{n} = (n_1, n_2, \cdots, n_N)$ and $\mathbf{p}^1$ must choose boxes sequentially to search for the gold coin, paying a cost $c_i > 0$ for each search of box, $i$, $1 \leqq i \leqq N$.

In a search of box $i$, the searcher withdraws a random number $W$ of coins whose distribution depends only on the number of coins in the box and the index $i$. For $0 \leqq w \leqq m \leqq n_i$, $f_i(w \mid m)$ denotes the probability that a search of box $i$ will withdraw $w$ coins when there are $m$ coins in the box. It is assumed that $f_i(0 \mid m) < 1$ for $1 \leqq m \leqq n_i$ and that $\sum_{w=0}^{m} f_i(w \mid m) = 1$, $0 \leqq w \leqq m \leqq n_i$. The searcher knows $\{f_i(w \mid m) : 0 \leqq w \leqq m \leqq n_i, 1 \leqq i \leqq N\}$. Furthermore, if the gold coin is in box $i$ and a search of box $i$ withdraws $w$ coins from the $m$ coins in the box, the probability that the gold coin is among those withdrawn is $w/m$. Thus $1 - w/m$ is the overlook probability for that search. The searcher learns the number of coins he withdraws at each stage. The problem is to minimize the total expected cost for finding the gold coin.

We assume, for $0 \leqq m < r \leqq n_i$, that

$$\text{(1)} \qquad \text{if } f_i(r - m \mid r) > 0 \text{ then } \sum_{w=0}^{m} w f_i(w \mid m) \leqq \sum_{w=0}^{r} w f_i(w \mid r), \quad 1 \leqq i \leqq N.$$

From our definition of $W$, $E_i(W \mid m) = \sum_{w=0}^{m} w f_i(w \mid m)$. Intuitively, (1) says that the expected batch size $E_i(W \mid m)$ is nonincreasing in $m$ as $m$ decreases.

THEOREM 1. *Suppose* (1) *holds whenever* $0 \leqq m \leqq r \leqq n_i$, $1 \leqq i \leqq N$. *Then an optimal search rule in the gold coin search problem is: If* $m_l$ *is the present number of coins in box* $l$, $1 \leqq l \leqq N$ *and* $\mathbf{p}$ *is the present location vector for the gold coin, search any box* $i$ *which achieves* $p_i E_i(W \mid m_i)/c_i m_i = \max_{1 \leq l \leq N} p_l E_l(W \mid m_l)/c_l m_l$ *where we include only those expressions for which* $m_l > 0$.

To prove the theorem, we note that if $W_{ij}$ is the random number of coins withdrawn during the $j$th search of $i$, the distribution of $\{W_{ij}\}$ is determined by the $\{f_i(w \mid m)\}$, and the overlook random variables are determined by $\{W_{ij}\}$. From (1), the S.M.C. is satisfied, so the B.B.K. rule is optimal.

(b) *The parametric adaptive search problem.* In this example we treat the adaptive search problem wherein the overlook random variables for box $i$ $\{\alpha_{ij}\}_{j=1}^{\infty}$ are i.i.d. given $\theta_i$, where $\theta_i$ is a fixed, unknown parameter. Moreover, we assume that prior to the beginning of the search nature chooses the parameters $\theta_1, \cdots, \theta_N$ independently from known prior distributions $F_1, \cdots, F_N$, respectively.

If $\varphi_{\theta,\sigma}(\mathbf{p}^1)$ denotes the (conditional) expected cost using policy $\sigma$ starting at state $\mathbf{p}^1$ when $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)$ is the true state of nature, $E\varphi_{\theta,\sigma}(\mathbf{p}^1)$ is the expected cost using $\sigma$, where $\boldsymbol{\theta}$ has distribution $\bigotimes_{i=1}^{N} F_i$. Since $\{\alpha_{ij}\}_{j=1}^{\infty}$ are i.i.d. given $\theta_i$, by computing the conditional posterior distribution of $\theta_i$ given $\alpha_{i1}, \cdots, \alpha_{i,m-1}$, we may find $E(\alpha_{im} \mid \alpha_{i1}, \cdots, \alpha_{i,m-1})$. To give an example for this search problem in which the S.M.C. holds, we restrict attention to a given box by dropping the subscript $i$, and let $\theta$ have the gamma density $g(\theta) = (\Gamma(a))^{-1} b^a \theta^{a-1} e^{-\theta b} I_{(0,\infty)}(\theta)$, $a > 0$, $b > 0$. Denote this distribution by $\mathscr{G}(a, b)$. Given $\theta$, let $\{\alpha_j\}_{j=1}^{\infty}$ be i.i.d. with density $f(\alpha \mid \theta) = \theta \alpha^{\theta-1} I_{[0,1]}(\alpha)$, $\theta > 0$. Then for $n \geqq 1$, the posterior distribution of $\theta$ given $\alpha_1, \cdots, \alpha_n$ is $\mathscr{G}(a + n, b - \sum_{j=1}^{n} \log \alpha_j)$. If $\lambda = a + n - 1$ and $\beta = b - \sum_{j=1}^{n-1} \log \alpha_j$ are fixed, one can easily show that (i) $E(1 - \alpha_n \mid \alpha_1, \cdots, \alpha_{n-1}) = E\{(1 + \beta^{-1} U)^{-1}\}$, where $U \in \mathscr{G}(\lambda, 1)$, and that (ii) $\psi(\alpha_n) = \alpha_n E(1 - \alpha_{n+1} \mid \alpha_1, \cdots, \alpha_n) = \alpha_n E\{(1 + (\beta - \log \alpha_n)^{-1} W)^{-1}\}$ is monotone increasing in

$\alpha_n \in [0, 1]$ provided $b \geqq 1$, where $W \in \mathscr{G}(\lambda + 1, 1)$. Since the quantity in (i) is at least $\psi(1)$, the S.M.C. is satisfied. Theorem 1 of Section 2 shows that $\varphi(\mathbf{p}) < \infty$.

(c) *The adaptive search problem with Dirichlet processes as nonparametric priors.* In this section we assume that the overlook random variables $\{\alpha_{ij}\}_{j=1}^{\infty}$ are i.i.d. according to a probability measure $P_i$ on $\mathscr{X}_i = [0, 1]$ which is unknown to the searcher. The measure $P_i$ is chosen and fixed at the beginning of the problem by nature according to a Dirichlet process (cf. Ferguson [4], [5]) with parameter $\lambda_i$, which the searcher does know. From Ferguson, $\lambda_i$ is a nonnull finite measure on $\mathscr{X}_i$, and $P_i$ is a random probability measure such that for each $k \geqq 1$ and every measurable partition $\{B_l\}_{l=1}^{k}$ of $\mathscr{X}_i$, the random vector $(P(B_1), \cdots, P(B_k))$ has a Dirichlet distribution with parameter $(\lambda_i(B_1), \cdots, \lambda_i(B_k))$; this is denoted by $P_i \in \mathscr{D}(\lambda_i)$. We assume $P_1, \cdots, P_N$ are chosen independently and let $\mathbf{P} = (P_1, \cdots, P_N)$.

If $\varphi_{\mathbf{P},\sigma}(\mathbf{p})$ denotes the (conditional) expected search cost using policy $\sigma$ when $\mathbf{P}$ is the true state of nature, $\varphi(\mathbf{p}) = \inf_\sigma \mathrm{E}(\varphi_{\mathbf{P},\sigma}(\mathbf{p}))$. A similar expression holds for the Sweat problem. Since $\{\alpha_{ij}\}_{j=1}^{\infty}$ are i.i.d. given $\mathbf{P}$, we can compute $E(\alpha_{im} \mid \alpha_{i1}, \cdots, \alpha_{i,m-1})$ if we can compute the conditional posterior distribution of a Dirichlet process given a sample from $P_i$. One of Ferguson's main results is that $P_i \mid \alpha_{i1}, \cdots, \alpha_{i,m-1} \in \mathscr{D}(\lambda_i + \sum_{j=1}^{m-1} \delta_{\alpha_{ij}})$, where $\delta_x$ is the probability measure with mass one at $x$. We are interested in finding conditions on $\lambda_i$ such that the S.M.C. is satisfied. We drop the subscript $i$ and assume that $\lambda(\{1\}) < 1$, so that the overlook will not be degenerate at 1.

Let $w$ denote $\lambda([0, 1])$ and $\nu = \int_0^1 x\, d\lambda(x)/w$ denote the prior mean. If a sample of size $n$ is taken of the Dirichlet process $P$ with observed values $t_1, \cdots, t_n$, the posterior mean given $t_1, \cdots, t_n$ is $\nu(t_1, \cdots, t_n) = \int_0^1 x\, d(\lambda + \sum_1^n \delta_{t_j})/(w + n) = (w\nu + \sum_1^n t_j)/(w + n)$. Thus for the S.M.C. to hold we need

$$t_{n+1}[1 - (w\nu + \textstyle\sum_1^n t_j + t_{n+1})/(w + n + 1)] \leqq 1 - (w\nu + \textstyle\sum_1^n t_j)/(w + n) \tag{2}$$

for almost all $t_j \in [0, 1]$ such that $\prod_1^n t_j > 0$, $1 \leqq j \leqq n + 1$, for all $n = 0, 1, 2, \cdots$. We have the following theorem.

THEOREM 2. *If* $(1 - \nu)w \geqq 1$, *then the* S.M.C. (2) *holds. If* 1 *is in the support of* $\lambda$, *then* (2) *holds if and only if* $(1 - \nu)w \geqq 1$.

PROOF. The proof is elementary and left to the reader. □

Unfortunately, it is not always true that $\varphi(\mathbf{p}^1) < \infty$, as simple examples show.

(d) *An adaptive search problem for which the* B.B.K. *search rule is not optimal.* Here we give an interesting example of a parametric adaptive search problem which does not satisfy the W.M.C. Set $N = 2$ and $c_1 = c_2 = 1$. Let the parameter space for $\theta_1$ be $\Theta_1 = \{\theta_1^0, \theta_1^1\}$ and let $f_1$ be the probability mass function of $F_1$ with $f_1(\theta_1^0) = \varepsilon$, $f_1(\theta_1^1) = 1 - \varepsilon$, where $0 < \varepsilon < 1$ will be chosen later. Let $\mu_1(\{\frac{1}{2}\} \mid \theta_1^0) = 1$ and $\mu_1(\{1 - \varepsilon\} \mid \theta_1^1) = 1$. For box 2, let $\Theta_2 = \{\theta_2^0\}$ and $\mu_2(\{\frac{1}{2}\} \mid \theta_2^0) = 1$. Thus the searcher already knows the value $\frac{1}{2}$ of the overlook random variable for box 2, and once he searches box 1 he learns the true value of the overlook probability there, and the problem then becomes one with known constant overlook probabilities.

We compute $E(1-\alpha_{11}) = \varepsilon(3/2-\varepsilon) \equiv \eta$. Choose $\varepsilon$, $(3-7^{\frac{1}{2}})/4 < \varepsilon < (3-5^{\frac{1}{2}})/4$, so that $\varepsilon$ is very close to $(3-7^{\frac{1}{2}})/4$. The W.M.C., $1 - E\alpha_{11} \geqq \alpha_{11}[1-E(\alpha_{12}\,|\,\alpha_{11})]$ a.s., fails to hold for $\alpha_{11} = \frac{1}{2}$. Let $p_1^1 = 1/(2\eta+1)$, $p_2^1 = 2\eta/(2\eta+1)$. According to the B.B.K. policy, the searcher may search either box 1 or box 2 first, since $p_1^1(1-E\alpha_{11}) = p_2^1(1-\alpha_2)$. Let $g''$ denote the version of the B.B.K. rule which searches box 2 first and $g'$ be the rule which first searches box 1 and then acts optimally. It is then strightforward to show $\varphi_{1,g'}(\mathbf{p}^1) < \varphi_{1,g''}(\mathbf{p}^1)$. It is easy to see from the computations that by changing $p_1^1$ slightly so that $p_1^1 < 1/(2\eta+1)$ but $p_1^1$ close to $1/(2\eta+1)$, the B.B.K. rule is no longer indifferent and must choose box 2 to search first. Since $g'$ will still be a better rule, the B.B.K. rule is not optimal.

**Acknowledgments.** I wish to thank my advisor, Professor Thomas S. Ferguson, for suggesting the original search problem that I worked on and for his guidance, suggestions and constant encouragement while I was writing my Ph. D. dissertation. Many thanks go to the referee for suggesting the much shorter proof of Theorem 1, Section 3, and his many invaluable comments.

## REFERENCES

[1] BELLMAN, RICHARD (1957). *Dynamic Programming*. Princeton Univ. Press.

[2] BLACK, WILLIAM L. (1965). Discrete sequential search. *Information and Control* **8** 159–162.

[3] BLACKWELL, DAVID (1962). Notes on dynamic programming. Unpublished notes. Univ. of California, Berkeley.

[4] BLACKWELL, DAVID (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 226–235.

[5] FERGUSON, THOMAS S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[6] FERGUSON, THOMAS S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

[7] HINDERER, K. (1970). *Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter*. Springer-Verlag, New York.

[8] KADANE, JOSEPH B. (1968). Discrete search and the Neyman-Pearson lemma. *J. Math. Anal. Appl.* **22** 156–171.

[9] ROSS, SHELDON M. (1969). A problem in optimal search and stop. *Operations Res.* **17** 984–992.

[10] STRAUCH, RALPH E. (1966). Negative dynamic programming. *Ann. Math. Statist.* **37** 871–889.

[11] SWEAT, CALVIN W. (1970). Sequential search with discounted income, the discount a function of the cell searched. *Ann. Math. Statist.* **41** 1446–1453.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712