Sequential Search on a Partition

Kevin A. Bryan*

March 4, 2020

Abstract

Consider costly sequential search for a prize located on a partition. Though the probabilities each box contains the prize are not independent, it is nonetheless possible to derive analytic results on the likelihood the prize is found. As in the case with independence, sequential search raises the likelihood of finding the prize compared to simultaneous search, but the underlying reason for the advantage is entirely different. Further, this likelihood depends on the full *distribution* of box sizes and not on a Pandora's Rule reservation value. The benefit of searching sequentially therefore depends critically on whether we are modeling a setting where the probability of finding a prize is truly independent across options.

1 Introduction

Consider a researcher scouring the literature for a half-remembered proof, a prospector trying to figure out where flecks of gold are draining from, or Edison searching through a set of potential filaments to create a rumored long-lasting incandescent lightbulb. In these cases, success probabilities across different options are negatively correlated: not finding the gold vein in one location raises the posterior probability the gold will be found in as-yet-unsearched spots. In these settings, what is the probability the "prize" will be found with optimal sequential search?

This problem is related to the famous search model of Weitzman [1979]. Let there be n boxes potentially containing a prize R, where the probability a given box i contains the prize is p_i, probabilities are independent, and each search costs c. In general, once the independence of p is dropped, optimal search is challenging to describe. As noted

*University of Toronto, Rotman School of Management. Correspondence: kevin.bryan@rotman.utoronto.ca.

by Weitzman [1979], if "reward distributions were not independent, the optimal search strategy could be very complicated...translating [informational externalities from correlated searches] into a simple search rule seems difficult except in the most elementary cases." Nonetheless, there is an important class of correlated sequential search which is both tractable and which has a particularly interesting implication for the value of being able to search sequentially rather than simultaneously: search on a partition.[1]

In this note, we show that the value of being able to search a partitioned set of potential prizes sequentially rather than simultaneously depends critically on the *distribution of prior probabilities* of each option being a success.[2] This distribution does not affect the search rule for simultaneous search on a partition, nor does it matter for sequential search with independent success probabilities for each option.[3] Further, though sequential search finds the prize with weakly higher probability than simultaneous search both on a partition and with independent success probabilities, the rationale is entirely different. Finally, the distribution of prior probabilities which maximize the likelihood of finding the prize under sequential partition search is generally not one where box sizes are similar.

2 Model

Let there be N "boxes" that may each contain a prize with value R. The cost of searching any box is c. The prior probability of finding a prize in box i is p_i, ordered from largest to smallest. A box either contains a prize or does not. You can keep at most one prize.[4] We will say boxes are "independent" if p_i is constant regardless of the realization of other searches. We will say boxes form a "partition" if at most one box

[1]Even minor extensions to the full Weitzman setting generally can be recast as a restless bandit problem, a class where analytical results have proved very difficult to come by. For example, Doval [2018] shows that a reservation value result maintains even if you can choose a box which has not been inspected, and take whatever that prize yields. We are able to show results both because we restrict to a binary prize setting and because of the particular negative correlation in payoff probabilities across boxes induced by a partition structure.

[2]Weitzman's full setting does not include this problem. Assume that each box generates an independent binary distribution F_i, such that the payoff is R with probability p_i and 0 otherwise. If a prize is found, then search immediately stopped. However, since probabilities are independent, *not* finding a prize in box i does not raise the probability the prize is in box j. Fundamentally, allowing this link is what drives our results in the present paper.

[3]Of course, the distribution matters for the probability of finding the prize under simultaneous search with independent probabilities since it affects the likelihood the prize is found more than once under a given search rule.

[4]Alternatively, the payoff is the maximum size of prizes found during the search.

contains a prize, $\sum_i p_i \leq 1$ and the posterior belief $\hat{p_i}$ is updated following search of other boxes according to Bayes' Rule.[5] Search may be simultaneous, in which case the set of boxes to be opened must be chosen all at once, or sequential, in which case the searcher can choose what order to open boxes, and can condition their strategy on what has been opened in the past.

Consider first three results which are either known in the literature, or straightforward implications of known results.

1) If p_i are *independent* and search is *simultaneous*, optimal search opens all boxes p_j such that $\prod_{i:i<j}(1-p_i)p_jR \geq c$, where boxes are ordered by size. We therefore find the prize with probability $1-\prod_{i\leq j*}(1-p_i)$ where $j*$ is the smallest box j such that $\prod_{i:i<j}(1-p_i)p_jR \geq c$. This is a direct implication of the marginal improvement algorithm in Chade and Smith [2006].[6]

2) If p_i are *independent* and search is *sequential*, optimal search opens boxes in order of p_i until the prize is found, or until there exists no box such that $p_iR \geq c$. We therefore find the prize with probability $1-\prod_{i:p_iR\geq c}(1-p_i)$ (Weitzman [1979]).[7] Note that sequential search opens weakly more boxes than simultaneous search.

3) If p_i *partitions* $[0,1]$, such that the single prize is in no box with probability $1-\sum_i p_i$, and in a given box with probability p_i, optimal *simultaneous* search opens every box such that $p_iR \geq c$. We therefore find the prize with probability $\sum_{i:p_iR\geq c} p_i$. Optimal simultaneous search with probabilities p_i finds the prize weakly more often than sequential search with independent probabilities, and the maximal number of boxes searched is identical. In both cases, failure to find a prize in a given box gives no information about the value of other boxes, because of independence and the fact that only one prize can be kept in the first case and because of the partition structure in the second case.

Let us now consider our primary case of interest, when p_i *partitions* $[0,1]$, and search is *sequential*. Recall that in the simplified version of Weitzman [1979] in case 2 above, the problem of which boxes to open, and in which order, depends on a reserva-

[5] We allow p to sum to less than one since the prior belief may include a belief that the prize does not exist with probability $1-\sum_i p_i$.

[6] The marginal improvement algorithm opens all boxes as long as the marginal value of opening a box exceeds the cost. The marginal improvement of the first box is p_1R, of the second box is $p_1R+(1-p_1)p_2R-p_1R=(1-p_1)p_2R$, of the third box is $p_1R+(1-p_1)p_2R+(1-p_1)(1-p_2)p_3R-p_1R-(1-p_1)p_2R=(1-p_1)(1-p_2)p_3R$, and so on.

[7] This is the famous Pandora's Rule with a particularly straightforward reservation value in the setting where boxes either have a prize or do not.

tion value which is based only on the properties of a particular box.[8] When we drop independence of probabilities, however, the decision to open is no longer a function only of the properties of that box. Nonetheless, some analytic results can be stated. The critical factor for optimal search, perhaps surprisingly, is the *distribution* of prize probabilities p.

Theorem 1. *Recall that the index i orders boxes by p_i from largest to smallest. Optimal sequential search on a partition:*

(i) opens boxes in order of p_i

(ii) is weakly more likely to find the prize as R increases, and weakly less likely as c increases or if p is replaced with a finer partition q.[9]

(iii) finds the prize weakly more often than under simultaneous search

(iv) opens all boxes until a prize is found as long as $\frac{p_i}{1-\sum_{k<i} p_k} \geq p_j, \forall i > j$, where j is the smallest box opened under simultaneous search.

(v) finds the prize with exactly the same probability as simultaneous search if $\frac{p_i}{1-\sum_{k<i} p_k} R \leq c, \forall i > j$, where again j is the lowest probability box opened under simultaneous search

(vi) finds the prize with a probability that depends on the full distribution of p. In particular, there exist partitions p where that probability increases when weight is shifted from smaller boxes to larger, and others where that probability increases when weight is shifted from larger boxes to smaller.

The proof is left for the appendix. Lemma 2 in the proof formally characterizes the implicit solution to the probability the prize is found as the property of a series of inequalities, and makes clear that no tighter analytic solution in terms of primitives exists.

To understand why more cannot be shown analytically, consider the following three numeric examples to understand the fundamental difficulty. Let there be two partitions $p = \{.4, .2, .15\}$ and $p' = \{.4, .25, .1\}$. Let $R = 1$.

Example 1. *Let search cost $c = .34$. If the first two boxes have been opened, the posterior probabilities that the final box contains the prize are $\frac{.15}{1-.6} = .375$ and $\frac{.1}{1-.65} = \frac{2}{7} \cong .28$. Therefore, the final box, should we have opened the first two boxes and not*

[8] And more broadly, the Weitzman reservation value even in the most general form does not depend on the distribution of probability mass in boxes with lower reservation value.

[9] Note that we specify these results in terms of the probability of finding the prize, not total boxes searched, since sequential search stops immediately if the prize is found.

found the prize, will be opened under p but not under p'. Working inductively, then, will we open the second biggest box conditional on opening the first? Note that the posterior probabilities the second biggest box contains the prize are $\frac{.2}{1-.4} = \frac{1}{3}$ and $\frac{.25}{1-.4} = \frac{5}{12} \cong .43$. But since under p we will open the third box conditional on not finding the prize in the second box, we need to consider this continuation value. The continuation value contribution to the payoff of opening the second box under p is the probability no prize is found in that box, times the posterior probability box 3 contains the prize, times the prize minus the search cost, or $\frac{2}{3}(.375 - .34) = \frac{7}{300}$. Therefore, the full payoff of opening box two under p is $\frac{1}{3} - .34 + \frac{7}{300} = \frac{1}{60}$, and under p' is $\frac{5}{12} - .34 = \frac{23}{300}$, hence the second box is opened in both partitions. Finally, the first box is worth opening in both partitions even if there were no continuation value. We have, then, that under p as many as three boxes will be opened, and the prize will be found with probability .75, but under p', up to two boxes will be opened, and the prize will be found with probability .65.

Example 2. *Now let the search cost be .39. In this case, the posterior probabilities are too low to open the third box in either partition. Therefore, we will open the second boxes if the posterior belief they contain a prize is at least .39, which only holds for p'. Again, it is worth opening the first box even if the continuation value after opening it is zero under both partitions. Hence, with this higher cost, under partition p we find the prize with probability .4, but under p' we find the prize with probability .65.*

Example 3. *Finally, consider the same partitions once more, but with the search cost .36. In this case, conditional on opening the first two boxes, the third box will be opened under p but not under p'. Working backward, will we open box two conditional on opening the largest boxes? Under p', the value of opening the second box is $\frac{5}{12} > .36$, so the second box is opened. However, under p, the value of opening the second box is the direct benefit $\frac{1}{3} - .36$ plus the continuation value $\frac{2}{3}(.375 - .36) = .01$, which sums to less than zero. Hence the second box will not be opened under p., and therefore neither will the third box. Finally, as before, the first boxes are always worth opening even if continuation values are zero, so under partition p we find the prize with probability .4, but under p' we find the prize with probability .65.*

These three examples make three features of sequential search on a partition clear. First, it is not straightforward to calculate whether "more equal" or "more varied" partitions make the prize more likely to be found. Second, the continuation value after opening a box plays a fundamental role. Third, the distribution of partitions can "cascade" forward or not.[10] In the first example, it wasn't worth opening the second box under p on its own, but it was given the option value of being able to open box three. In the third example, we still would have opened box three had we opened box two without success, but even given this option value, it wasn't worth opening box two. In general, whether we open any given box depends on the value of opening that box in addition to option value if the prize is not found, where that option value depends on the value of future boxes including *their* option value, and so on. The decision to open or not depends on more than just the parameters of a particular box. Indeed, the proof shows that to answer the question "with what probability will the prize be found" for a partition with N boxes where $p_i > 0$ requires checking up to $N!$ inequalities.

Note also the difference compared to sequential search with independent probabilities, or simultaneous search on a partition. In those cases, we search only those boxes where $p_i R \geq c$, and hence need only know the prior probability of success. In all three examples, for both p and p', only the first box is searched, and the prize is found with probability .4.

To clarify part (vi) of the theorem, consider the partitions that lead the prize to found with probability at least $\bar{p}$. Naturally, we can shift all weight to the largest partition. In this case, the prize is found if $\bar{p}R \geq c$, just by searching the largest partition. What about *marginal* shifts in weight, however? Consider starting with a partition p. When we move a small amount of weight from a smaller partition to a larger partition, the posterior probability of finding a prize when searching the smaller partition p_i falls. This is not a trivial claim: although the direct prior probability p_i falls, the summed probability of all larger partitions rises by an equal amount, hence the increase from the prior to the *posterior* belief the prize in i conditional on not finding the prize before reaching i rises. Nonetheless, that indirect effect of moving weight w is smaller than the direct effect: $\frac{\partial}{\partial w}\frac{p_i+w}{1-\sum_{j<i}p_j-w} > 0$.

We therefore face a tradeoff. By moving weight to larger partitions, we increase the probability a prize is found when we search those partitions. However, we do so at the cost of lowering the likelihood we will search the now-smaller partitions if the prize is not found. In the numerical example above, moving weight from the third box to the second box both directly makes the benefit of searching the second box higher, and lowers the posterior probability of finding a prize in, and therefore of searching,

[10]Even in the most general Weitzman [1979] setting, though reservation values depend on the distribution of probability mass, they are constant in any rearrangement of mass that leaves the total mass in boxes with lower reservation values constant.

the third box. That tradeoff increased the probabliity the prize was found in Examples 2 and 3, but not in Example 1.

This tradeoff suggests there may be multiple qualitatively-different distributions that will find a prize with probability at least $\bar{p}$ for any $\bar{p} \leq 1$, and indeed there are.

Theorem 2. *Let there be reward R and search cost c. For any $\bar{p} \in (0, 1]$, let N^* be the smallest N such that $\sum_{i=0}^{N} \frac{c}{R}(1 - \frac{c}{R})^N \geq \bar{p}$. Let $\bar{p}R \geq c$.*

(i) The partition $\{\bar{p}, 0, 0....0\}$ and the partition $\{\frac{c}{R}, \frac{c}{R}(1 - \frac{c}{R}), \frac{c}{R}(1 - \frac{c}{R})^2, \frac{c}{R}(1 - \frac{c}{R})^3 ... \frac{c}{R}(1 - \frac{c}{R})^{N^*}, 0, ...0\}$ both find the prize with at least probability $\bar{p}$.

(ii) Further, there exists w>0 such that shifting weight w from any box to any other box in either distribution ensures that the prize is not found with at least probability $\bar{p}$.

(iii) If $\sum_{i=0}^{N^*} \frac{c}{R}(1 - \frac{c}{R})^{N^*} \geq \bar{p}$ holds with equality, then there exists no distribution $p' \neq p$ such that $p'_1 \leq p_1$, $\frac{p_i}{p_{i-1}} \leq \frac{p'_i}{p'_{i-1}}, \forall i$ and the prize is found with probability at least $\bar{p}$.

That is, we can find the prize with any given probability either by making the problem trivial - the prize is found with probability $\bar{p}$ by opening one box of size $\bar{p}$ - or by spreading out the distribution such that the posterior after every failed search is just high enough to continue searching. Note also that the latter distribution maximizes the advantage of sequential search compared to simultaneous search. Since under simultaneous search, only partitions with $p_i R \geq c$ are searched, only the first box is opened, and the price is found with probability $\frac{c}{R}$. On the other hand, with sequential search, the prize is found with probability $\bar{p}$, where $\bar{p}$ can be any probability up to and including 1. Likewise, the distribution $\{\bar{p}, 0, 0....0\}$ minimizes the benefit of sequential search, since in both sequential and simultaneous search only the first box is opened.

3 Discussion

Generically, sequential search is better than simultaneous search for two reasons. First, onces a prize is found, other boxes do not need to be wastefully serached. Second, when a prize is not found, the posterior probability other boxes contain the prize can rise. The first factor entirely drives the difference in behavior when p_i are independent: simultaneous search finds the prize less frequently than sequential search because the simultaneous searcher does not want to waste effort "finding the prize twice". However, it is the second factor which entirely drives the difference in behavior when p_i is a partition: it is not possible to find the prize twice, but failed searches increase the posterior belief that remaining boxes contain a prize, hence the option value of continued search becomes relevant. That is, the reason sequential search finds the prize more often is entirely different in the independent and partition setups.

The general search problem with correlated probabilities remains very difficult to characterize analytically. Nonetheless, settings where partitioned rather than fully independent search is more realistic are common. We have shown that this setting has a (relatively) straightforward characterization. As we have seen in Theorem 2, by manipulating the distribution of p, we can, for any reward and search cost, induce sequential searchers to find the prize with arbitrarily high probability while not changing the likelihood a simultaneous searcher, or a sequential searcher with independent p, finds the prize. Since the use of sequential search on a partition is so different from independent search, these results may prove particularly useful to applied theorists in settings where option value rather than avoided repetition is the driving motivation.

4 Appendix: Proofs

4.1 Proof of Theorem 1

(i) Assume that $p_i > p_j$, yet p_j is searched first. Whether or not p_i will be searched under the proposed optimal strategy, by replacing p_j with p_i, total expected search costs weakly fall, and the total expected reward strictly rises, hence searcher payoff strictly increases.

Let us describe precisely the optimal search strategy given (i) before returning to the remaining proofs.

Assume that p_N is the final box which will be searched. It is ex-post optimal to search p_N if and only if $\frac{p_N}{1-\sum_{i<N} p_i} R - c \geq 0$. The first term is the posterior probability that the prize is in partition N conditional on it not being found in prior partitions. If that condition holds, it is optimal to search partition p_{N-1} if and only if

$$\frac{p_{N-1}}{1-\sum_{i<(N-1)} p_i} R - c + \left(1 - \frac{p_{N-1}}{1-\sum_{i<(N-1)} p_i}\right) V_{N-1} \geq 0$$

where V is the continuation utility. Applying induction and rearranging terms, we have that

Lemma 1. *p_N is the final box searched, and hence $\sum_{i \leq N} p_N$ is the probability the prize* *is found, if* $\forall j \in 1, ..., N$,

$$\frac{\sum_{k:j\leq k\leq N} p_k}{1-\sum_{i<j} p_i} R \geq [(N+1-j) - \frac{\sum_{k:j\leq k<N}(N-k)p_k}{1-\sum_{i<j} p_i}]c$$

and these inequalities do not hold for all $j \in 1, ..., N+1$.

Reading from left to right, this inequality says that, at every posterior i from the initial prior before we search p_1 until the final potential search before we search p_N, the posterior probability of finding the prize times the prize value exceeds the total costs incurred. The cost term accounts for the fact that, when considered from the posterior beliefs after searching from 1 to $i-1$, the search cost c will definitely be paid once, will be paid twice if we do not find a prize searching p_i, will be paid thrice if we do not find the prize in p_i or p_{i+1}, and so on. Note, as in the example in the main text, that calculating the exact maximal number of boxes searched, and hence probability of finding the prize, involves the interaction between each of the p_i probabilities, and hence there is no simple "reservation value" type of calculation to be made. Note also that we only search until p_N if the value function is positive at *every* continuation from the ex-ante prior until the posterior right before searching p_N.

(ii) Trivial from the characterization in Lemma 2.

(iii) Under simultaneous search on a partition, all boxes are searched such that the prior expected payoff $p_i R$ exceeds the search cost c. Under sequential search, the value of opening any given box is weakly higher than under simultaneous both because posterior beliefs $\frac{p_j}{1-\sum_{k<j} p_k}$ are weakly higher than the prior p_j, and because the continuation value is weakly bigger than zero. Therefore, any box opened under simultaneous search with a given search cost must also be opened under sequential search.

(iv) Since j is opened under simultaneous search, we know that $p_j R \geq c$. By (iii), we know all boxes weakly larger than p_j are opened under sequential search. By assumption, the posterior belief that any box smaller than j contains the prize, conditional on having opened all larger boxes, is at least p_j. Since continuation values are weakly larger than zero, the expected reward from opening every box therefore is larger than the search cost.

(v) By (iii), all boxes weakly larger than j are opened. If $\frac{p_i}{1-\sum_{k<i} p_k} R \leq c, \forall i > j$, the final box is not opened, hence there is no continuation value and the second smallest box is not opened, and so on. Hence the number of boxes is identical to under simultaneous search.

(vi) We have a direct proof of this statement in the three examples in the main text.

4.2 Proof of Theorem 2

(i) That the first partition finds the prize is trivial. For the second partition, note that it is optimal to open any box, regardless of continuation value, if $\frac{p_i}{1-\sum_{j<i} p_j} \geq \frac{c}{R}$. If that equation holds with equality for all boxes opened in a proposed search strategy, there is no continuation value after any search. That equation holding with equality implies that $p_1 = \frac{c}{R}$ and $\frac{p_i}{1-\sum_{k<i} p_k} = \frac{p_j}{1-\sum_{k<j} p_k}, \forall i, j$.

Combining those equalities, it can be shown by induction that $p_i = \frac{c}{R}(1 - \frac{c}{R})^{i-1}$ is the distribution p such that the posterior probability at every i is just sufficient to make continued search worthwhile.

(ii) In the first case, moving weight from p_1 to any other box j gives a posterior probability $\frac{w}{1-p_1-w}$, which is zero in the limit. Since there is no continuation value after opening box j, and since the posterior limits to 0 as w goes to zero, with sufficiently small zero, box j will not be opened. Hence the prize will be found with likelihood lower than $\bar{p}$.

In the second case, the distribution p was chosen so that the expected value of opening every box is exactly zero. Moving weight from box j to larger box i therefore ensures j will not be opened, and hence the prize is found with likelihood lower than $\bar{p}$. Moving weight from j to smaller box i creates continuation value, but that continuation value is strictly lower than the direct value of keeping the weight in j (multiple search costs c are paid to reach that weight). Therefore, j will not be opened, and again the prize with found with likelihood less than $\bar{p}$.

Note that $\sum_{i=0}^{\infty} x(1-x)^N = 1, \forall x \in (0, 1)$, so this distribution will eventually sum to $\bar{p}$ for any $\bar{p} \in (0, 1)$.

(iii) If the condition holds with equality, optimal search finds the prize with precisely probability $\bar{p}$ and continuation value following failed search is exactly zero following any failed search. If $p'_1 \leq p_1$, then $\frac{p_i}{p_{i-1}} \leq \frac{p'_i}{p'_{i-1}}, \forall i$ and $p' \neq p$ implies that $p'_i < p_i$ for at least one box, and that there is no continuation value following any search. Hence by construction of p_i, it is not optimal to open box i.

References

Hector Chade and Lones Smith. Simultaneous search. *Econometrica*, 74(5):1293–1307, 2006.

Laura Doval. Whether or not to open pandora's box. *Journal of Economic Theory*, 175:127–158, 2018.

Martin L Weitzman. Optimal search for the best alternative. *Econometrica*, pages 641–654, 1979.