



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 262

Optimal Search from Multiple Distributions with Infinite Horizon

Jean-Michel Benkert, Igor Letina and Georg Nöldeke

Revised version, December 2017

Optimal Search from Multiple Distributions with Infinite Horizon

Jean-Michel Benkert^a, Igor Letina^{b,*}, Georg Nöldeke^c

^a*Department of Economics and UBS International Center of Economics in Society,
University of Zurich, Switzerland*

^b*Department of Economics, University of Bern, Switzerland*

^c*Faculty of Business and Economics, University of Basel, Switzerland*

Abstract

With infinite horizon, optimal rules for sequential search from a known distribution feature a constant reservation value that is independent of whether recall of past options is possible. We extend this result to the case when there are multiple distributions to choose from: it is optimal to sample from the same distribution in every period and to continue searching until a constant reservation value is reached.

JEL classification: D83

Keywords: Optimal Search, Search Intensity, Infinite Horizon, Recall

*Corresponding author. Igor Letina, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland. Tel.: +41 31 631 80 76. Email: igor.letina@vwi.unibe.ch

1. Introduction

In the standard optimal search problem (e.g. Lippman and McCall, 1976), the choice is to either stop searching and consume the best option available or to continue the costly search. Continuation yields a draw of a single observation from some known distribution. With infinite horizon, the optimal search rule is simple and independent of the possibility of recall: continue searching until a constant reservation value, uniquely determined by the distribution, has been reached.

A natural extension of the standard problem is to allow the searcher to also choose the search intensity, modelled as the number of simultaneous observations drawn from the known distribution.¹ With infinite horizon, Morgan (1983, Proposition 1) shows that with no recall it is optimal to search with constant intensity until stopping. With full recall, Morgan (1983, Proposition 5) only establishes that the optimal intensity is weakly decreasing, leaving open the possibility that the searcher might reduce intensity after a favorable draw. In an interesting application to delegated R&D, Poblete and Spulber (2017, Lemma 3) show optimality of constant intensity with full recall while assuming existence, uniqueness and reservation value strategies. We strengthen these results and show that the optimal intensity is constant — *regardless* of the possibility of recall and without restricting the class of admissible search rules.

The choice of search intensity is a special case of a more general problem in

¹There are alternative ways to model search intensity. For example, Karni and Schwartz (1977a) model it as the amount of time between two consecutive search attempts.

which the searcher chooses a distribution from which to draw the observation. For example, the searcher could be choosing the riskiness of search. Similarly, no recall and full recall are special cases of a more general problem in which recall is stochastic as in Landsberger and Peled (1977). We incorporate both of these generalizations in an otherwise standard search model with infinite horizon and show that it is optimal to sample from the same distribution in every period and to continue searching until the constant reservation value associated with this distribution has been achieved.

As we will discuss in more detail later, our result is a very natural one in light of the related literature. First, once the infinite horizon of the search problem is taken into account, our optimal search rule corresponds to the one that Weitzman (1979) has identified as optimal for a search problem with full recall and finite horizon. Second, the irrelevance of recall obtains for the same reason as in the related results for search problems with infinite horizon in DeGroot (1970, p. 335), Lippman and McCall (1976, p. 169), and Landsberger and Peled (1977, Theorem 2), namely that the option of recall is never exercised when full recall is possible.

2. Model

While search is ongoing, a searcher decides in each period $t = 0, 1, \dots$ whether to sample from one of n available distributions (take the action $a_t \in A = \{1, \dots, n\}$) or to stop ($a_t = 0$). Stopping yields a payoff $x_t \in X$ in the current period, where x_t is the option the searcher has in hand at the beginning of period t and $X \subset \mathbb{R}_{++}$ is a finite set. Once search has been stopped, no further decisions can be taken and no further payoffs accrue. If

the searcher samples from distribution a_t , the state variable x_t transitions to $x_{t+1} \in X$ with probability $p(x_{t+1}|x_t, a_t)$, where x_{t+1} is the option the searcher has in hand at the beginning of period $t + 1$. The cost of sampling from distribution a is $c(a) > 0$. The per-period discount factor is $\delta \in (0, 1]$. The searcher's problem is to maximize her expected discounted payoff (for every initial condition $x_0 \in X$) by choosing a search rule $\mu : X \rightarrow A \cup \{0\}$ that specifies for each state whether search should be continued by sampling from distribution a ($\mu(x) = a$) or be stopped ($\mu(x) = 0$).²

The above model can be embedded into a Markov decision process framework (Bertsekas, 1995) by (i) appending a terminal state $x = 0$ that is reached with probability one whenever the stopping action $a = 0$ is taken and (ii) supposing that the only available action (at zero cost) in the terminal state is $a = 0$. With discounting ($\delta < 1$) standard results (e.g. Bertsekas, 1995, Chapter 1.2) ensure that the Bellman equations

$$v^*(x) = \max \left\{ x, \max_{a \in A} \left\{ \delta \sum_{y \in X} v^*(y) p(y|x, a) - c(a) \right\} \right\} \quad (1)$$

have a unique solution $v^* : X \rightarrow \mathbb{R}$ and that a search rule μ^* is optimal iff it satisfies

$$\mu^*(x) = \begin{cases} 0 & \Rightarrow v^*(x) = x \\ a' \in A & \Rightarrow v^*(x) = \delta \sum_{y \in X} v^*(y) p(y|x, a') - c(a'). \end{cases} \quad (2)$$

In particular, an optimal search rule exists. The same conclusions hold without discounting ($\delta = 1$) because the Markov decision process formulation

²The restriction to such stationary search rules is without loss of generality; see the sources cited in the next paragraph.

of our search problem is a special case of the stochastic shortest path problem analysed in Bertsekas and Tsitsiklis (1991).³

We impose additional structure on the conditional probability distributions $p(\cdot|x, a)$ that accommodates the familiar cases of search with no or full recall but also allows for more general specifications.

Assumption 1. For all $a \in A$ there exists a probability distribution $q(\cdot|a)$ on X such that $p(y|x, a) = q(y|a)$ holds for all $y > x \in X$.

If the condition $p(y|x, a) = q(y|a)$ holds for all x and y in X the only option available to the searcher in period $t + 1$ after choosing a in period t is the realized draw from the distribution $q(\cdot|a)$, so that there is no recall. With perfect recall, the option x_t remains available in period $t + 1$ after a draw from $q(\cdot|a)$ has been taken, which corresponds to Assumption 1 holding with $p(x|x, a) = \sum_{y \leq x} q(y|a)$, and $p(y|x, a) = 0$ for $y < x$. Stochastic recall as in Landsberger and Peled (1977) is obtained by taking $p(\cdot|x, a)$ to be convex combinations of the probability distributions describing the no-recall and the full-recall case. Assumption 1 covers this case while allowing for more general specifications.

³Assumption 1 in Bertsekas and Tsitsiklis (1991) is satisfied: First, there exists an absorbing, cost- (and benefit)-free state (the terminal state 0). Second, there exists a proper stationary policy (choose the terminal action in each state). Third, improper stationary rules (policies for which there is a strictly positive probability that the terminal state is never reached) result in infinite expected cost because $c(a) > 0$ for all $a \in A$. Assumption 2 in Bertsekas and Tsitsiklis (1991) holds because A is finite. Our claim then follows from Proposition 2 in Bertsekas and Tsitsiklis (1991).

3. Optimal Search Rules

Define for each action $a \in A$ the reservation value $s(a)$ as the unique solution (Step 1, proof of Proposition 1) to the equation

$$\delta \sum_{y \in X} \max\{y, s(a)\} q(y|a) - s(a) = c(a). \quad (3)$$

When there is only one distribution to sample from, it is well-known (DeGroot, 1970; Lippman and McCall, 1976) that both in the no-recall and the full-recall case the optimal rule is to continue searching until the current option x_t exceeds the reservation value $s(a)$. The same kind of reservation value rule is optimal under Assumption 1 when searching from multiple distributions. Moreover, the searcher optimally samples from the same distribution a^* with the highest reservation value until the reservation value $s(a^*)$ of this distribution has been achieved:

Proposition 1. *Let Assumption 1 hold and let $a^* \in \arg \max s(a)$. The unique solution to the Bellman equations (1) is given by $v^*(x) = \max\{x, s(a^*)\}$ and the rule*

$$\mu^*(x) = \begin{cases} 0 & \text{if } x > s(a^*) \\ a^* & \text{if } x \leq s(a^*) \end{cases} \quad (4)$$

is optimal.

We provide a direct and straightforward proof in the appendix. To obtain intuition, consider the problem from Weitzman (1979). There, a searcher called Pandora faces a finite number of closed boxes with an uncertain reward hidden in each box. For each box there is a cost of opening it. The problem is to determine the order in which to open the boxes, and when to stop searching

and accept the highest reward sampled so far. The optimal rule (Pandora's Rule) assigns a reservation value to each box, depending only on the properties of that box, and then specifies to open the boxes in descending order by reservation value until the highest sampled reward exceeds the reservation value of every closed box.

With full recall, our search problem is like Pandora's problem when there are infinite copies of each of n different types of boxes. The optimal search rule identified in Proposition 1 is nothing but Pandora's Rule applied to this scenario, namely to keep opening boxes of the type with the highest reservation value until the most recent draw exceeds this reservation value. Since search only continues if all previously sampled rewards are below the reservation value of the best type of box and the option to open another such box is always available, this rule never uses the possibility to recall an earlier reward. As the optimality of a rule cannot be affected by eliminating options that the rule never exercises, the result in Proposition 1 holds regardless of whether recall is possible or not. The fact that Assumption 1 suffices for the result can be understood as a generalization of the observation in Weitzman (1979) that the optimal search rule is not affected by rearranging the probability distribution for rewards below the reservation value.

4. Discussion

Our result can be extended in several directions. First, provided that the reservation value $s(a)$ is well-defined for all $a \in A$ and that the problem $\max_{a \in A} s(a)$ has a solution, the arguments proving Proposition 1 go through without any substantial modification. In particular, under standard regularity

conditions (Bertsekas and Tsitsiklis, 1991; James and Collins, 2006) the set of distributions A can be compact or countably infinite rather than finite and the support of the probability distributions $q(\cdot|a)$ can be compact as in Poblete and Spulber (2017).

Second, it is only for the convenience of modelling search as a simple Markov decision process that we have assumed transition probabilities to depend only on the best option currently in hand and that whether or not such an option is recalled in the next period is exogenous. While not fitting into our formal framework, more elaborate models of recall, such as the ones in Karni and Schwartz (1977a), Karni and Schwartz (1977b) and Ikuta (1988), in which the probability of a successful recall depends on the number of periods since a draw has been taken, or the one in Saito (1998), in which recall is costly, cannot upset our conclusion that with an infinite horizon it is optimal to search until the reservation value $s(a^*)$ is exceeded and stop then.⁴ In particular, recalling a past option can only be optimal “off the equilibrium path” when such an option should have been accepted at the time it had become available.

Third, as noted in Banks and Sundaram (1992, footnote 3) and discussed in more detail in Bergemann and Välimäki (2001, Section 4), the results obtained in Weitzman (1979) can be viewed as establishing the optimality of a Gittins index policy in a multi-armed bandit problem in which a switching cost has to be paid on the first trial with a given bandit. The same interpretation is applicable in our setting. In particular, under assumptions analogous to

⁴Of course, our result is not applicable if the horizon is finite or the cost of sampling is increasing over time – which are the cases that these papers focus on.

the ones maintained in Bergemann and Välimäki (2001, Theorem 2), the appropriately defined Gittins index policy is optimal in a multi-armed bandit problem with switching costs in which there is an infinite number of each of n different types of arms, with all the arms of the same type being ex ante identical. Just as in our model, the possibility of recall is irrelevant in such an environment with stationary bandits because an optimal policy never requires the agent to return to a bandit that has previously been abandoned (Banks and Sundaram, 1992; Bergemann and Välimäki, 2001).

Acknowledgments

We thank Julia Grünseis and an anonymous referee for suggestions.

Appendix: Proof of Proposition 1

STEP 1: The function $g : A \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$g(s, a) = \delta \sum_{y \in X} \max\{y, s\} q(y|a) - s \quad (5)$$

is (strictly) decreasing in s for all $a \in A$ (if $s < \max\{X\}$). This follows from rewriting

$$g(s, a) = \delta \sum_{y \in X} \max\{y - s, 0\} q(y|a) - (1 - \delta)s$$

and noting (DeGroot, 1970, Sec. 11.8) that $\sum_{y \in X} \max\{y - s, 0\} q(y|a)$ is (strictly) decreasing in s (whenever it is strictly positive).

Observing that (3) is equivalent to

$$g(s(a), a) = c(a) \quad (6)$$

and that $c(a) > 0$ for all $a \in A$, it is immediate from the above monotonicity properties that (3) can have at most one solution $s(a)$. Existence of a solution follows upon observing that $g(s, a)$ is continuous, satisfies $g(\max\{X\}, a) = 0$, and $\lim_{g \rightarrow -\infty} g = \infty$.

STEP 2: Let $v^*(x) = \max\{x, s(a^*)\}$.

Consider $x \leq s(a^*)$. We then have for all $a \in A$:

$$\begin{aligned}
\delta \sum_{y \in X} v^*(y)p(y|x, a) - c(a) &= \delta \sum_{y \in X} \max\{y, s(a^*)\}p(y|x, a) - c(a) \\
&= \delta \sum_{y \in X} \max\{y, s(a^*)\}q(y|a) - c(a) \\
&= g(s(a^*), a) - c(a) + s(a^*) \\
&= g(s(a^*), a) - g(s(a), a) + s(a^*) \\
&\leq s(a^*),
\end{aligned}$$

where the first equality is from the specification of v^* , the second uses Assumption 1 to infer that for $y > s^*(a)$ the condition $x \leq s^*(a)$ implies $p(y|x, a) = q(y|x)$, the third is from the definition of g in (5), and the fourth is from (6). The inequality in the last line follows from Step 1 and $s(a^*) \geq s(a)$.

It holds as an equality for $a = a^*$. Therefore

$$s(a^*) = \delta \sum_{y \in X} v^*(y)p(y|x, a^*) - c(a^*) = \max \left\{ x, \max_{a \in A} \left\{ \delta \sum_{y \in X} v^*(y)p(y|x, a) - c(a) \right\} \right\} \tag{7}$$

holds for all $x \leq s(a^*)$.

Consider $x > s(a^*)$. We then have for all $a \in A$:

$$\delta \sum_{y \in X} v^*(y)p(y|x, a) - c(a) = \delta \sum_{y \in X} \max\{y, s(a^*)\}p(y|x, a) - c(a)$$

$$\begin{aligned}
&\leq \delta \sum_{y \in X} \max\{y, x\} p(y|x, a) - c(a) \\
&= \delta \sum_{y \in X} \max\{y, x\} q(y|a) - c(a) \\
&= g(x, a) - c(a) + x \\
&= g(x, a) - g(s(a), a) + x \\
&\leq x,
\end{aligned}$$

where the equality in the first line again uses $v^*(x) = \max\{x, s(a^*)\}$, the inequality in the second line uses $x > s^*(a)$, the equality in the third line is from Assumption 1, and the remaining lines are obtained in the same way as in the case $x \leq s(a^*)$, using that $x > s(a^*)$ implies $x > s(a)$ for all $a \in A$ to obtain the final inequality. Therefore,

$$x = \max \left\{ x, \max_{a \in A} \left\{ \delta \sum_{y \in X} v^*(y) p(y|x, a) - c(a) \right\} \right\} \quad (8)$$

holds for all $x > s(a^*)$.

Combining (7) and (8), it is immediate that $v^*(x) = \max\{x, s(a^*)\}$ solves the Bellman equations (1) and that $\mu^*(x)$ as defined in (4) satisfies the optimality conditions (2).

References

- BANKS, J. S. AND R. K. SUNDARAM (1992): “Denumerable-Armed Bandits,” *Econometrica*, 60, 1071–1096.
- BERGEMANN, D. AND J. VÄLIMÄKI (2001): “Stationary Multi-Choice Bandit Problems,” *Journal of Economic Dynamics & Control*, 25, 1585–1594.

- BERTSEKAS, D. P. (1995): *Dynamic Programming and Optimal Control*, vol. 2, Athena Scientific.
- BERTSEKAS, D. P. AND J. N. TSITSIKLIS (1991): “An Analysis of Stochastic Shortest Path Problems,” *Mathematics of Operations Research*, 16, 580–595.
- DEGROOT, M. H. (1970): *Optimal Statistical Decisions*, Hoboken, New Jersey: John Wiley & Sons.
- IKUTA, S. (1988): “Optimal Stopping Problem with Uncertain Recall,” *Journal of the Operations Research Society of Japan*, 31, 145–171.
- JAMES, H. W. AND E. COLLINS (2006): “An Analysis of Transient Markov Decision Processes,” *Journal of Applied Probability*, 43, 603–621.
- KARNI, E. AND A. SCHWARTZ (1977a): “Search Theory: The Case of Search with Uncertain Recall,” *Journal of Economic Theory*, 16, 38–52.
- (1977b): “Two Theorems on Optimal Stopping with Backward Solicitation,” *Journal of Applied Probability*, 14, 869–875.
- LANDSBERGER, M. AND D. PELED (1977): “Duration of Offers, Price Structure, and the Gain from Search,” *Journal of Economic Theory*, 16, 17–37.
- LIPPMAN, S. A. AND J. MCCALL (1976): “The Economics of Job Search: A Survey,” *Economic Inquiry*, 14, 155–189.
- MORGAN, P. B. (1983): “Search and Optimal Sample Size,” *Review of Economic Studies*, 50, 659–675.

- POBLETE, J. AND D. SPULBER (2017): “Managing Innovation: Optimal Incentive Contracts for Delegated R&D with Double Moral Hazard,” *European Economic Review*, 95, 38–61.
- SAITO, T. (1998): “Optimal Stopping Problem with Controlled Recall,” *Probability in the Engineering and Informational Sciences*, 12, 91–108.
- WEITZMAN, M. L. (1979): “Optimal Search for the Best Alternative,” *Econometrica*, 47, 641–654.