Optimal Stopping with a Capacity Constraint: Generalizing Shepp's Urn Scheme

Amin Dehghanian

Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran

Jeffrey P. Kharoufeh*

Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

We formulate an optimal stopping problem for a variant of Shepp's urn model in which it is possible to sample more than one item at each stage. Using a Markov decision process model, we establish monotonicity of the optimal value function and show that the optimal policy is a monotone threshold policy that prescribes either not sampling, or sampling the maximum number of items permitted. A special case exhibits convexity and submodularity, but these properties do not hold in general.

Keywords: Optimal stopping, sampling policy, urn models, Markov decision process

---

1. Introduction

Motivated by an optimal stopping problem introduced by Breiman [1] and expanded upon by Chow and Robbins [2], Teicher and Wolfowitz [3], and Dvoretzky [4], Shepp [5] examined a related discrete-time stopping problem that can be described as follows. Consider an urn initially consisting of p perfect and i imperfect items, where p and i are finite, positive integers that are known *a priori*. A single item is sampled randomly from the urn sequentially (without replacement) until some appropriate stopping time. The item type (perfect or imperfect) is revealed only after it is drawn; for each imperfect item drawn, a reward of $+1$ is received, and for each perfect

*Corresponding author: Jeffrey P. Kharoufeh. Email: jkharouf@pitt.edu.

item drawn, a reward (or cost) of -1 is incurred. At each step, the decision maker may decide not to draw any item at all. Shepp addressed the question of when to stop sampling from the urn to ensure that the total expected return is positive. He showed that the optimal policy for this stopping problem has a threshold structure and then asymptotically identified the boundary of the continuation region by employing a Brownian bridge approximation to the problem. Specifically, for a given number of imperfect items i, there is a threshold $\beta(i)$ such that if $p \leq \beta(i)$, it is optimal to draw from the urn; if $p > \beta(i)$, it is optimal not to draw. (A similar optimal threshold-type policy holds when p and i are interchanged.) Additionally, Shepp showed that $\lim_{i\to\infty}(\beta(i) - i)(2i)^{-0.5} = \alpha \approx 0.83992$. The optimal stopping problem that we examine here can be viewed as a generalization of Shepp's urn scheme in that we allow sampling multiple items at each step and use a different cost/reward structure of which Shepp's is a special case.

Optimal stopping problems similar to Shepp's have been analyzed over the past several decades with the models of Boyce [6, 7] being most relevant to our work here. Boyce [6] investigated a bond-selling problem for which there exist short-term price fluctuations and mid-term price information. Using a pinned Brownian motion process, he considered the problem of finding the optimal time to sell a bond. Specifically, he derived the optimal selling time under the somewhat restrictive condition that the Brownian motion is pinned to a Gaussian distribution at a certain time. Subsequently, he considered Shepp's urn model as a finite approximation of the pinned Brownian motion. This approximation relaxes the Gaussian restriction and is more computationally appealing. Boyce further refined this approximation by generalizing Shepp's urn model to a random urn model in which the total number of items in the urn is deterministic and known, while the number of imperfect items is known only through a probability distribution. A recursive algorithm was devised to compute the optimal value function and optimal policy. Boyce [7] established structural properties of the optimal value function, including the optimality of a threshold-type policy which had been established by Shepp [5], by a recursive method. Boyce developed an efficient algorithm to compute the threshold values and corroborated Shepp's asymptotic result that $\alpha \approx 0.83992$. We devise and analyze an optimal stopping problem similar to Shepp's urn model but from a stochastic optimization point of view. Some of the key differences are as follows:

- The urn model we analyze allows for sampling multiple items at once rather than a single item. Shepp's model can be viewed as a special case of ours in which the maximum number of drawn items is $c = 1$;

- The cost/reward structure of our model is less restrictive in that a reward of r is received for each sampled imperfect item, but no cost is incurred for drawing a perfect item;
- This sequential decision problem is addressed via a Markov decision process (MDP) model to establish important structural properties of the optimal value function and optimal policy. We prove that it is optimal to either (1) sample none of the items (thereby stopping the process), or (2) sample the maximum number possible. Hence, for any state, if it is optimal to sample, then it is optimal to sample the maximum number allowable (*full inspection policy*). This optimal policy is one of monotone thresholds. By contrast, Shepp provided an asymptotic continuation region by employing a continuous-time approximation and a scaled process;
- We examine in detail the case with $c = 1$ (Shepp's model with a modified cost/reward structure) and establish the convexity and submodularity of the optimal value function. Interestingly, these properties do not necessarily hold for $c > 1$ as demonstrated via a counterexample.

The remainder of the paper is organized as follows. Section 2 provides the detailed problem formulation and our MDP model. Section 3 establishes properties of the optimal value function and optimal policy. Section 4 provides convexity and submodularity results for the special case $c = 1$. Finally, in Section 5, we illustrate the behavior of the optimal value function by way of a numerical example and demonstrate that the convexity and submodularity results do not necessarily hold when $c > 1$.

2. Problem Description and MDP Formulation

Consider an urn initially containing I imperfect and P perfect items, respectively ($I, P \in \mathbb{N}$). At each decision epoch, the current number of imperfect and perfect items within the urn is known, and one must decide the number of items to randomly sample from the urn (without replacement). This number can be selected from the set $\{0, 1, \ldots, c\}$, where c is a finite, positive integer. However, if fewer than c items are contained in the urn, then we are limited by the number in the urn. The type of each item (perfect or imperfect) is revealed only when it is drawn. If at any decision epoch it is optimal to sample 0 items, the process is stopped, no cost is incurred and no reward is gained; however, if at least one item is drawn, then a fixed reward (cost) of -1 is incurred, irrespective of the number of items drawn.

For each imperfect item drawn, a reward of r ($r > 0$) is received, and for each perfect item drawn, no reward is earned and no cost is incurred. This sequential sampling process continues until the decision is to sample none of the items from the urn with the objective of maximizing the total expected profit accrued until the process is stopped.

Next, we introduce notation and formulate this sequential decision problem using an MDP model. Let $\mathbb{Z}_+ = \{0, 1, \ldots\}$ be the set of nonnegative integers and $\mathbb{N} = \{1, 2, \ldots\}$ the set of natural numbers. For any $n \in \mathbb{N}$, C_m^n denotes the number of ways to select m items from a collection of n items, and we adopt the convention that $C_n^{-1} = 0$. We also assume sums of the form $\sum_0^{-1}(\cdot) = 0$. We consider an infinite horizon MDP model in which the decision epochs are the elements of $\mathbb{N}$. The time interval between any two consecutive decision epochs, n and $n+1$, is called the nth period. The state of the MDP model is an ordered pair (i, p), where i and p denote the current number of imperfect and perfect items in the urn, respectively, at the start of the current period. Let $V(i, p)$ be the maximum total expected profit accrued until stopping when starting in state (i, p) over all non-anticipative policies (i.e., those that depend on the past and present but not the future), and similarly let $V[(i, p), j]$ be the maximum total expected profit over all such policies, provided that $j \geq 1$ items are drawn in the current period. We pause here to note that $V(i, p)$ exists and is finite, as the state and action spaces are finite and bounded, the one-stage costs/rewards are finite and bounded, and the number of decision epochs before stopping, under each non-anticipative policy, is bounded. Hence, the value function of each non-anticipative policy is well-defined, finite and bounded. Furthermore, the set of non-anticipative policies is finite, so $V(i, p)$ exists and is finite for each state (i, p). Theorems 7.1.7 and 7.1.9 of [8] establish the existence of a stationary, deterministic optimal policy that satisfies Bellman's optimality equation, which is described in what follows.

For any state (i, p), let $S_{ip} := \{1, \ldots, \min\{c, i + p\}\}$. Therefore, for each state (i, p), Bellman's optimality equation is

$$V(i, p) = \max \left\{ \max_{j \in S_{ip}} \{V[(i, p), j]\}, 0 \right\}, \tag{1}$$

where, for each $j \in S_{ip}$,

$$V[(i, p), j] = -1 + \sum_{k=0}^{j} \frac{C_k^i C_{j-k}^p}{C_j^{i+p}} \big(kr + V(i - k, p - j + k)\big). \tag{2}$$

The index k in (2) denotes the number of imperfect items drawn in the current period. For any state (i, p), the least that can be drawn is zero, and the most that can be drawn is $j \in S_{ip}$. The zero term in equation (1) represents the reward for stopping the process.

In our model, drawing no item is equivalent to stopping the process. Alternatively, one can devise a model in which the action space includes the option of sampling 0 items (at no cost) and allowing the process to continue. However, the revised model has an optimal value function and optimal policy that coincide with our model. To see this, couple any policy in the revised model with an admissible policy to the original model that follows the same sequence of actions except that the action of sampling no item is eliminated wherever it appears. The value functions of these two policies coincide; hence, the optimal policy of the revised model can be obtained by only searching over the set of policies that never choose the action 'sample 0 items and continue to the next decision epoch'.

Remark 1. *If $c = 1$ and $r = 2$, the optimality equation is as follows:*

$$\begin{aligned} V(i,p) &= \max\left\{-1 + \frac{p}{i+p}V(i,p-1) + \frac{i}{i+p}\left(2 + V(i-1,p)\right), 0\right\} \\ &= \max\left\{\frac{p}{i+p}\left(-1 + V(i,p-1)\right) + \frac{i}{i+p}\left(1 + V(i-1,p)\right), 0\right\}, \quad (3) \end{aligned}$$

which is identical to the one derived by Boyce [7] for Shepp's urn model. In this sense, our model can be viewed as a generalization of Shepp's model.

3. Structural Properties

In this section, we establish important properties of the optimal value function and characterize the optimal policy. However, before doing so, we require the following lemma.

Lemma 1. *For each state $(i, p-1) \in \mathbb{Z}_+^2$,*

$$V(i+1, p-1) - V(i,p) \leq r. \quad (4)$$

Proof. We will prove the result by induction on $i + p$. It is clear that inequality (4) holds for $i + p = 1$. For the induction hypothesis, assume the same holds for $i + p = n > 1$; we proceed to show that it holds for

$i + p = n + 1$. Note that,

$$
\begin{aligned}
\sum_{k=0}^{j} & C_k^i C_{j-k}^p \big(kr + V(i-k, p-j+k)\big) \\
&= \sum_{k=1}^{j+1} C_{k-1}^i C_{j-k+1}^p \big((k-1)r + V(i-k+1, p-j+k-1)\big) && \text{(5a)} \\
&= \sum_{k=1}^{j+1} C_{k-1}^i C_{j-k+1}^p \big(kr + V(i-k+1, p-j+k-1)\big) && \text{(5b)} \\
&\quad - \sum_{k=1}^{j+1} C_{k-1}^i C_{j-k+1}^p r \\
&= \sum_{k=1}^{j+1} C_{k-1}^i C_{j-k+1}^p \big(kr + V(i-k+1, p-j+k-1)\big) - C_j^{i+p} r, && \text{(5c)}
\end{aligned}
$$

where (5a) and (5c) follow from a simple index transformation and Vandermonde's identity, respectively. Therefore,

$$
\begin{aligned}
V&[(i+1, p-1), j] - V[(i,p), j] \\
&= \frac{1}{C_j^{i+p}} \Bigg\{ \sum_{k=0}^{j} C_k^{i+1} C_{j-k}^{p-1} \big(kr + V(i-k+1, p-j+k-1)\big) \\
&\qquad\qquad - \sum_{k=0}^{j} C_k^i C_{j-k}^p \big(kr + V(i-k, p-j+k)\big) \Bigg\} && \text{(6a)} \\
&= \frac{1}{C_j^{i+p}} \Bigg\{ \sum_{k=0}^{j+1} C_k^{i+1} C_{j-k}^{p-1} \big(kr + V(i-k+1, p-j+k-1)\big) \\
&\qquad - \sum_{k=0}^{j+1} C_{k-1}^i C_{j-k+1}^p \big(kr + V(i-k+1, p-j+k-1)\big) - C_j^{i+p} r \Bigg\} && \text{(6b)} \\
&= \frac{\sum_{k=0}^{j+1} \big(C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^i C_{j-k+1}^p\big)\big(kr + V(i-k+1, p-j+k-1)\big)}{C_j^{i+p}} + r, && \text{(6c)}
\end{aligned}
$$

where (6b) follows from (5a)–(5c) and the fact that $C_{-1}^n := 0$ for all $n \in \mathbb{Z}_+$. It is sufficient to show that the summation on the right-hand side (rhs) of (6c) is nonpositive. Note that $C_k^{i+1} C_{j-k}^{p-1} \geq C_{k-1}^i C_{j-k+1}^p$ if and only if $k \leq \underline{k}$, where

$$
\underline{k} := \left\lfloor \frac{(i+1)(j+1)}{i+p+1} \right\rfloor .
$$

From Vandermonde's identity, it follows that

$$\sum_{k=0}^{\underline{k}} C_k^{i+1} C_{j-k}^{p-1} + \sum_{k=\underline{k}+1}^{j+1} C_k^{i+1} C_{j-k}^{p-1} = \sum_{k=0}^{\underline{k}} C_{k-1}^{i} C_{j-k+1}^{p} + \sum_{k=\underline{k}+1}^{j+1} C_{k-1}^{i} C_{j-k+1}^{p} = C_j^{i+p}.$$

Therefore,

$$\sum_{k=0}^{\underline{k}} C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p} = \sum_{k=\underline{k}+1}^{j+1} C_{k-1}^{i} C_{j-k+1}^{p} - C_k^{i+1} C_{j-k}^{p-1}. \tag{7}$$

By the induction hypothesis, for each $k \in \{0, \dots, \underline{k}\}$,

$$kr + V(i-k+1, p-j+k-1) \leq \underline{k}r + V(i-\underline{k}+1, p-j+\underline{k}-1), \tag{8a}$$

and for each $k \in \{\underline{k}+1, \dots, j+1\}$,

$$kr + V(i-k+1, p-j+k-1) \geq \underline{k}r + V(i-\underline{k}+1, p-j+\underline{k}-1). \tag{8b}$$

Hence,

$$\begin{aligned}
&\sum_{k=0}^{j+1} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p})\big(kr + V(i-k+1, p-j+k-1)\big) \\
&= \sum_{k=0}^{\underline{k}} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p})\big(kr + V(i-k+1, p-j+k-1)\big) \qquad (9a)\\
&\quad + \sum_{k=\underline{k}+1}^{j+1} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p})\big(kr + V(i-k+1, p-j+k-1)\big) \\
&\leq \sum_{k=0}^{\underline{k}} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p})\big(\underline{k}r + V(i-\underline{k}+1, p-j+\underline{k}-1)\big) \qquad (9b)\\
&\quad + \sum_{k=\underline{k}+1}^{j+1} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p})\big(\underline{k}r + V(i-\underline{k}+1, p-j+\underline{k}-1)\big) \\
&= \big(\underline{k}r + V(i-\underline{k}+1, p-j+\underline{k}-1)\big) \sum_{k=0}^{j+1} (C_k^{i+1} C_{j-k}^{p-1} - C_{k-1}^{i} C_{j-k+1}^{p}) \qquad (9c)\\
&= 0, \qquad (9d)
\end{aligned}$$

where (9b) follows from (8a)–(8b) and the definition of $\underline{k}$, and (9d) follows from Vandermonde's identity. We conclude from (9a)-(9d) that the summation on the rhs of (6c) is nonpositive, which completes the proof. ■

Starting in state (i, p) and choosing $j \geq 1$ items to draw in this period, $kr + V(i-k, p-j+k) - 1$ is the total expected profit under the realization that k of the j drawn items are imperfect. Lemma 1 implies that the function $kr + V(i-k, p-j+k)$ is nondecreasing in k; therefore, it is more advantageous to detect imperfect items earlier rather than later. The next theorem establishes monotonicity of the value function.

Theorem 1. *For each* $(i,p) \in \mathbb{Z}_+^2$,

(i) $V(i+1,p) \geq V(i,p)$,

(ii) $V(i,p+1) \leq V(i,p)$.

Proof. For part (i), we again employ induction on $i+p$. The inequality in Theorem 1(i) certainly holds for $i+p=0$. For the induction hypothesis, assume that it holds for $i+p=n>1$. Setting $\overline{k} := \lceil (i+1)j/(i+p+1) \rceil$, we see that

$$\frac{C_k^{i+1} C_{j-k}^p}{C_j^{i+p+1}} \geq \frac{C_k^i C_{j-k}^p}{C_j^{i+p}}$$

if and only if $k \geq \overline{k}$. Therefore,

$$\begin{aligned} &V[(i+1,p),j] - V[(i,p),j] \\ &\quad = \frac{1}{C_j^{i+p} C_j^{i+p+1}} \Bigg\{ \sum_{k=0}^{j} C_j^{i+p} C_k^{i+1} C_{j-k}^p \big(kr + V(i-k+1, p-j+k)\big) \\ &\qquad\qquad - \sum_{k=0}^{j} C_j^{i+p+1} C_k^i C_{j-k}^p \big(kr + V(i-k, p-j+k)\big) \Bigg\}, \end{aligned} \tag{10a}$$

and the bracketed expression in (10a) can be expressed as

$$\begin{aligned} &\sum_{k=0}^{\overline{k}-1} C_j^{i+p} C_k^{i+1} C_{j-k}^p \big(V(i-k+1, p-j+k) - V(i-k, p-j+k)\big) \\ &+ \sum_{k=\overline{k}}^{j} C_j^{i+p+1} C_k^i C_{j-k}^p \big(V(i-k+1, p-j+k) - V(i-k, p-j+k)\big) \\ &- \sum_{k=0}^{\overline{k}-1} \Big(C_j^{i+p+1} C_k^i C_{j-k}^p - C_j^{i+p} C_k^{i+1} C_{j-k}^p \Big) \big(kr + V(i-k, p-j+k)\big) \\ &+ \sum_{k=\overline{k}}^{j} \Big(C_j^{i+p} C_k^{i+1} C_{j-k}^p - C_j^{i+p+1} C_k^i C_{j-k}^p \Big) \big(kr + V(i-k+1, p-j+k)\big). \end{aligned} \tag{10b}$$

By the induction hypothesis, the first and second summations of (10b) are nonnegative. Hence, it suffices to show that the third summation does not exceed the fourth. By Lemma 1, it follows that for each $k \in \{0, \ldots, \overline{k}-1\}$,

$$kr + V(i-k, p-j+k) \leq \overline{k}r + V(i-\overline{k}, p-j+\overline{k}), \tag{11}$$

and for each $k \in \{\overline{k}, \ldots, j\}$,

$$kr + V(i-k+1, p-j+k) \geq \overline{k}r + V(i-\overline{k}+1, p-j+\overline{k}). \tag{12}$$

By reasoning similar to that used to establish (7), it can be shown that

$$\begin{aligned}\sum_{k=0}^{\overline{k}-1} C_j^{i+p+1} C_k^i C_{j-k}^p - C_j^{i+p} C_k^{i+1} C_{j-k}^p = \sum_{k=\overline{k}}^{j} & C_j^{i+p} C_k^{i+1} C_{j-k}^p \\ & - C_j^{i+p+1} C_k^i C_{j-k}^p. \end{aligned} \tag{13}$$

Therefore,

$$\begin{aligned}
&\sum_{k=\overline{k}}^{j} (C_j^{i+p} C_k^{i+1} C_{j-k}^p - C_j^{i+p+1} C_k^i C_{j-k}^p)\big(kr + V(i-k+1, p-j+k)\big) \\
&\quad - \sum_{k=0}^{\overline{k}-1} (C_j^{i+p+1} C_k^i C_{j-k}^p - C_j^{i+p} C_k^{i+1} C_{j-k}^p)\big(kr + V(i-k, p-j+k)\big) \\
&\geq \sum_{k=\overline{k}}^{j} (C_j^{i+p} C_k^{i+1} C_{j-k}^p - C_j^{i+p+1} C_k^i C_{j-k}^p)\big(\overline{k}r + V(i-\overline{k}+1, p-j+\overline{k})\big) \qquad (14a) \\
&\quad - \sum_{k=0}^{\overline{k}-1} (C_j^{i+p+1} C_k^i C_{j-k}^p - C_j^{i+p} C_k^{i+1} C_{j-k}^p)\big(\overline{k}r + V(i-\overline{k}, p-j+\overline{k})\big) \\
&= \sum_{k=\overline{k}}^{j} (C_j^{i+p} C_k^{i+1} C_{j-k}^p - C_j^{i+p+1} C_k^i C_{j-k}^p) \qquad (14b) \\
&\qquad \times \big(V(i-\overline{k}+1, p-j+\overline{k}) - V(i-\overline{k}, p-j+\overline{k})\big) \\
&\geq 0, \qquad (14c)
\end{aligned}$$

where inequality (14a) follows from (11)–(12) and the definition of $\overline{k}$, (14b) follows from (13), and (14c) follows from the induction hypothesis and the definition of $\overline{k}$. This completes the proof of part (i).

For part (ii), let $\underline{k} := \lfloor ij/(i+p+1) \rfloor$ and note that $\frac{C_k^i C_{j-k}^{p+1}}{C_j^{i+p+1}} \geq \frac{C_k^i C_{j-k}^p}{C_j^{i+p}}$ if and only if $k \leq \underline{k}$. The remainder of the proof mirrors that of part (i). ∎

Remark 2. *For Shepp's urn scheme, Boyce [7] showed that the last item drawn is imperfect. Theorem 1 (ii) implies that our last draw must include at least one imperfect item.*

Theorem 2. *For each* (i,p) *and* $1 \leq j \leq c-1$, $V[(i,p),j+1] \geq V[(i,p),j]$.

Proof. There is an integer $\overline{k}$ such that given i, p, and j,

$$\frac{C_k^i C_{j+1-k}^p}{C_{j+1}^{i+p}} \geq \frac{C_k^i C_{j-k}^p}{C_j^{i+p}}$$

if and only if $k \geq \overline{k}$. Next, we note that

$$\begin{aligned} &V[(i,p),j+1] - V[(i,p),j] \\ &\quad = \frac{1}{C_j^{i+p} C_{j+1}^{i+p}} \Bigg\{ \sum_{k=0}^{j+1} C_j^{i+p} C_k^i C_{j+1-k}^p \big(kr + V(i-k, p-j-1+k)\big) \\ &\qquad\qquad - \sum_{k=0}^{j} C_{j+1}^{i+p} C_k^i C_{j-k}^p \big(kr + V(i-k, p-j+k)\big) \Bigg\}, \end{aligned} \tag{15a}$$

and the bracketed expression on the rhs of (15a) can be written as

$$\begin{aligned} &\sum_{k=0}^{\overline{k}-1} C_j^{i+p} C_k^i C_{j+1-k}^p \big(V(i-k, p-j-1+k) - V(i-k, p-j+k)\big) \\ &+ \sum_{k=\overline{k}}^{j+1} C_{j+1}^{i+p} C_k^i C_{j+1-k}^p \big(V(i-k, p-j-1+k) - V(i-k, p-j+k)\big) \\ &- \sum_{k=0}^{\overline{k}-1} \Big(C_{j+1}^{i+p} C_k^i C_{j-k}^p - C_j^{i+p} C_k^i C_{j+1-k}^p\Big) \big(kr + V(i-k, p-j+k)\big) \\ &+ \sum_{k=\overline{k}}^{j+1} \Big(C_j^{i+p} C_k^i C_{j+1-k}^p - C_{j+1}^{i+p} C_k^i C_{j-k}^p\Big) \big(kr + V(i-k, p-j-1+k)\big). \end{aligned} \tag{15b}$$

The proof is completed in a manner similar to that of Theorem 1. ∎

Corollary 1. *For each state* $(i, p) \in \mathbb{Z}_+^2$*, the optimal action is to either (a) draw no item at all, or (b) draw the maximum possible number of items,* $\min\{c, i + p\}$*. That is, the optimality equations can be expressed as*

$$V(i, p) = \max\{V[(i, p), \min\{c, i + p\}], 0\}, \quad (i, p) \in \mathbb{Z}_+^2. \tag{16}$$

If $i + p \leq c$, the one-stage expected profit of drawing $i + p$ items is equal to $-1 + ir/(i + p)$, so the optimality equation can simply be rewritten as $V(i, p) = \max\{-1 + ir/(i + p), 0\}$. Moreover, $V(0, p) = 0$ for all $p \in \mathbb{Z}_+$. Using this initialization, we can use the recursive formula (16) to compute $V(i, p)$ for each $(i, p) \in \mathbb{Z}_+^2$. Indeed, we adopt such an approach to present a numerical illustration of our model in Section 4.

The next theorem characterizes the optimal policy, which is of threshold type. Let $a^*(i, p)$ be the optimal action in state (i, p).

Theorem 3. *The optimal policy is one of monotone thresholds. That is, for a given* $i \in \mathbb{Z}_+$*, there exists a number* $p^*(i)$ *such that*

$$a^*(i, p) = \begin{cases} \min\{i + p, c\}, & p < p^*(i), \\ 0, & p \geq p^*(i), \end{cases} \tag{17}$$

and the threshold $p^*(i)$ *is nondecreasing in* i*. Furthermore, for a given* $p \in \mathbb{Z}_+$*, there exists a number* $i^*(p)$ *such that*

$$a^*(i, p) = \begin{cases} \min\{i + p, c\}, & i \geq i^*(p), \\ 0, & i < i^*(p), \end{cases} \tag{18}$$

and the threshold $i^*(p)$ *is nondecreasing in* p.

Proof. To establish (17), let

$$p^*(i) := \min\{p \in \mathbb{Z}_+ : V(i, p) \leq 0\}. \tag{19}$$

If the set defined in (19) is empty, then put $p^*(i) = +\infty$. Note that $V(i, p) \leq 0$ is identical to $V[(i, p), j] \leq 0$ for all $j \geq 1$. Now $p < p^*(i)$ implies that $V(i, p) > 0$; hence, it is optimal to draw $\min\{i + p, c\}$ items in state (i, p) by equation (16). On the other hand, if $p \geq p^*(i)$, then

$$V(i, p) \leq V(i, p^*(i)) \leq 0, \tag{20}$$

where the first inequality follows from Theorem 1 (*ii*), and the second inequality follows from the definition of $p^*(i)$. Thus, it is optimal in state (i, p)

not to draw any items by (16). It remains to show that the threshold $p^*(i)$ is nondecreasing in i. By Theorem 1 and the definition of $p^*(i)$, we see that

$$V(i+1, p^*(i)-1) \geq V(i, p^*(i)-1) > 0.$$

This implies that $p^*(i+1) \geq p^*(i)$. The proof for (18) is similar. ■

The optimal policy of Theorem 3 can be described as *full inspection* since, if it is optimal to draw, then it is optimal to draw the maximum allowable number of items. It differs from the policy described by Shepp in that we do not explicitly determine $p^*(i)$ or $i^*(p)$, whereas Shepp characterizes the continuation region using a continuous approximation. It may be instructive in future work to examine the asymptotic behavior of $p^*(i)$ as $i \to \infty$ in our generalized setting. In Section 4, we prove additional properties of the optimal value function for the special case when $c = 1$.

4. Convexity and Submodularity

Here we establish that the optimal value function is convex and submodular when $c = 1$. This case includes, but is not limited to, Shepp's urn scheme for which convexity was established by Chen and Hwang [9]. While similar results exist for a different model (see [10]), we provide a counterexample in Section 5 demonstrating that these properties do not necessarily hold if $c > 1$ for our model setting. Next, we present the definition of submodularity, and then Theorem 4, which is our main result in this section.

Definition 1. *[8] A function $f : \mathbb{Z}_+^2 \to \mathbb{R}$ is said to be submodular if for each $(n, m) \in \mathbb{Z}_+^2$,*

$$f(m+1, n+1) + f(m, n) \leq f(m+1, n) + f(m, n+1).$$

Theorem 4. *If $c = 1$, then the following inequalities hold:*

(i) $V(i+1, p) + V(i-1, p) \geq 2V(i, p)$,

(ii) $V(i, p+1) + V(i, p-1) \geq 2V(i, p)$,

(iii) $V(i+1, p) - V(i, p) \leq V(i+1, p-1) - V(i, p-1)$.

Proof. To prove parts (i)–(iii), we will use induction on $i+p$. Note that (i)–(iii) certainly hold for $i + p = 1$. For the induction hypothesis, assume that they each hold for $i + p = n > 1$.

First, we establish part (i) for $i + p + 1$. If it is optimal to draw no item in state (i, p) (i.e., if $V(i, p) = 0$), Theorem 4 (i) follows from the nonnegativity of the optimal value function. Otherwise, it follows from the optimality equation (1) that:

$$(i+p+1)V(i+1,p) \geq -(i+p+1) + (i+1)\big(r + V(i,p)\big) \\ + pV(i+1,p-1), \quad (21a)$$

$$(i+p+1)V(i-1,p) \geq -(i+p-1) + (i-1)\big(r + V(i-2,p)\big) \\ + pV(i-1,p-1) + 2V(i-1,p), \quad (21b)$$

$$2(i+p+1)V(i,p) = -2(i+p) + 2i\big(r + V(i-1,p)\big) \\ + 2pV(i,p-1) + 2V(i,p). \quad (21c)$$

Hence,

$$\begin{aligned}(i+p+1)&\left[V(i+1,p) + V(i-1,p) - 2V(i,p)\right] \\ &= (i-1)\left[V(i,p) + V(i-2,p) - 2V(i-1,p)\right] \\ &\qquad + p\left[V(i+1,p-1) + V(i-1,p-1) - 2V(i,p-1)\right] \\ &\geq 0,\end{aligned}$$

where the inequality follows from the induction hypothesis. This completes the proof of part (i).

Next, to prove part (ii), write inequalities for $(i+p+1)V(i,p+1)$, $(i+p+1)V(i,p-1)$, and $2(i+p+1)V(i,p)$ similar to those for (21a)–(21c), and follow the same steps.

Finally, we show that part (iii) holds for $i+p+1$. If $V(i+1,p) = 0$, then the desired result immediately follows from Theorem 1 (i). Otherwise, by Theorem 1 (i), it is sufficient to show that the following expression is nonpositive:

$$\begin{aligned}V&[(i+1,p),1] - V(i,p) - [V[(i+1,p-1),1] - V(i,p-1)] \\ &= \frac{(i+1)r + p\left[V(i+1,p-1) - V(i,p)\right]}{i+p+1} \\ &\qquad - \frac{(i+1)r + (p-1)\big(V(i+1,p-2) - V(i,p-1)\big)}{i+p} \qquad (22a) \\ &= \frac{1}{(i+p)(i+p+1)}\left\{-(i+1)r + p(i+p)\left[V(i+1,p-1) - V(i,p)\right]\right. \\ &\qquad \left. -(p-1)(i+p+1)\big(V(i+1,p-2) - V(i,p-1)\big)\right\}. \qquad (22b)\end{aligned}$$

The bracketed expression in (22b) can be rearranged as follows:

$$p(i+p)\,[V(i+1,p-1) - V(i,p) - V(i+1,p-2) + V(i,p-1)] \\ + (i+1)\,[V(i+1,p-2) - V(i,p-1) - r]\,. \quad (22c)$$

The expression on the first line of (22c) is nonpositive since

$$V(i+1,p-1) - V(i+1,p-2) \le V(i,p-1) - V(i,p-2) \quad (23a)$$
$$\le V(i,p) - V(i,p-1), \quad (23b)$$

where (23a) follows from the induction hypothesis, and (23b) follows from Theorem 4 (*ii*). The expression on the second line of (22c) is nonpositive by Lemma 1; therefore, (22c) is nonpositive and the proof is complete. ■

5. Numerical Example

The purpose of this section is to provide a counterexample demonstrating that convexity and submodularity do not necessarily hold if $c > 1$. Table 1 provides the optimal value function for a problem instance with $r = 2$ and $c = 2$. The following inequalities demonstrate that convexity/concavity and submodularity/supermodularity of the optimal value function cannot be definitively asserted when $c > 1$:

$$V(3,1) - V(2,1) > V(2,1) - V(1,1) > V(4,1) - V(3,1), \quad (24)$$
$$V(1,1) - V(1,0) > V(1,3) - V(1,2) > V(1,2) - V(1,1), \quad (25)$$
$$V(1,1) + V(2,2) > V(1,2) + V(2,1), \quad (26)$$
$$V(1,2) + V(2,3) < V(1,3) + V(2,2). \quad (27)$$

Specifically, (24)–(25) show that $V(i,p)$ is neither convex nor concave in i and p. Similarly, (26)–(27) show that $V(i,p)$ is neither submodular nor supermodular.

Acknowledgements. The authors acknowledge, with gratitude, an anonymous referee whose constructive comments have improved a prior version of the manuscript.

[1] L. Breiman, Stopping rule problems, Applied Combinatorial Mathematics (1964) 284–319.

[2] Y. S. Chow, H. Robbins, On optimal stopping rules for S_n/n, Illinois Journal of Mathematics 9 (1965) 444–454.

Table 1: The optimal value function $V(i,p)$ when $r=2$ and $c=2$.

	p									
i	0	1	2	3	4	5	6	7	8	9
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	1.00	1.00	0.67	0.50	0.20	0.00	0.00	0.00	0.00	0.00
2	3.00	2.33	2.17	1.70	1.47	1.05	0.79	0.50	0.29	0.05
3	4.00	4.00	3.40	3.20	2.69	2.43	1.98	1.70	1.32	1.05
4	6.00	5.20	5.07	4.46	4.23	3.68	3.41	2.92	2.63	2.19
5	7.00	7.00	6.29	6.11	5.49	5.25	4.68	4.40	3.89	3.59
6	9.00	8.14	8.04	7.35	7.14	6.52	6.26	5.68	5.40	4.86
7	10.00	10.00	9.22	9.07	8.39	8.16	7.54	7.27	6.69	6.39
8	12.00	11.11	11.02	10.28	10.09	9.42	9.18	8.55	8.28	7.69
9	13.00	13.00	12.18	12.05	11.32	11.11	10.45	10.20	9.57	9.29

[3] H. Teicher, J. Wolfowitz, Existence of optimal stopping rules for linear and quadratic rewards, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 5 (4) (1966) 361–368.

[4] A. Dvoretzky, Existence and properties of certain optimal stopping rules, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 1967, pp. 441–452.

[5] L. A. Shepp, Explicit solutions to some problems of optimal stopping, The Annals of Mathematical Statistics 40 (3) (1969) 993–1010.

[6] W. M. Boyce, Stopping rules for selling bonds, The Bell Journal of Economics and Management Science (1970) 27–53.

[7] W. M. Boyce, On a simple optimal stopping problem, Discrete Mathematics 5 (4) (1973) 297–312.

[8] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, 2014.

[9] R. W. Chen, F. K. Hwang, On the values of an (m,p) urn, Congr. Numer. 41 (1984) 75–84.

[10] R. W. Chen, A. Zame, A. M. Odlyzko, L. A. Shepp, An optimal acceptance policy for an urn scheme, SIAM Journal on Discrete Mathematics 11 (2) (1998) 183–195.