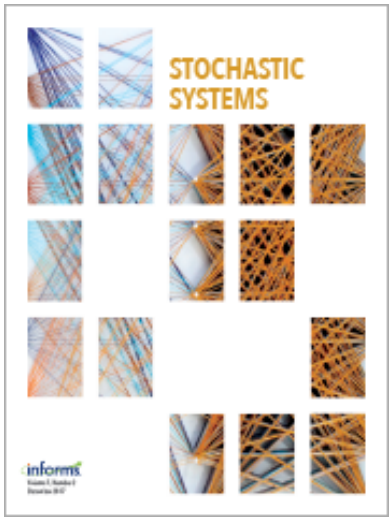


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Uniformly Bounded Regret in the Multisecretary Problem

Alessandro Arlotto, Itai Gurvich

To cite this article:
Alessandro Arlotto, Itai Gurvich (2019) Uniformly Bounded Regret in the Multisecretary Problem. Stochastic Systems 9(3):231-260. https://doi.org/10.1287/stsy.2018.0028

http://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 9, No. 3, September 2019, pp. 231–260
ISSN 1946-5238 (online)

Uniformly Bounded Regret in the Multisecretary Problem

Alessandro Arlotto,[a] **Itai Gurvich**[b]

[a] The Fuqua School of Business, Duke University, Durham, North Carolina 27708; [b] Cornell Tech and School of Operations Research and Information Engineering, Cornell University, New York, New York 10044

Contact: alessandro.arlotto@duke.edu, http://orcid.org/0000-0003-2705-9814 (AA); gurvich@cornell.edu, http://orcid.org/0000-0001-9746-7755 (IG)

Received: January 1, 2018
Revised: June 1, 2018
Accepted: October 29, 2018
Published Online in Articles in Advance: August 30, 2019

https://doi.org/10.1287/stsy.2018.0028

Abstract. In the secretary problem of Cayley [Cayley A (1875) Mathematical questions with their solutions. *Ed. Times* 23:18–19.] and Moser [Moser L (1956) On a problem of Cayley. *Scripta Mathematica* 22(3/4):289–292.], n nonnegative, independent, random variables with common distribution are sequentially presented to a decision maker who decides when to stop and collect the most recent realization. The goal is to maximize the expected value of the collected element. In the k-choice variant, the decision maker is allowed to make $k \le n$ selections to maximize the expected total value of the selected elements. Assuming that the values are drawn from a known distribution with finite support, we prove that the best regret—the expected gap between the optimal online policy and its offline counterpart in which all n values are made visible at time 0—is uniformly bounded in the number of candidates n and the budget k. Our proof is constructive: we develop an adaptive budget-ratio policy that achieves this performance. The policy selects or skips values depending on where the ratio of the residual budget to the remaining time stands relative to multiple thresholds that correspond to middle points of the distribution. We also prove that being adaptive is crucial: in general, the minimal regret among nonadaptive policies grows like the square root of n. The difference is the value of adaptiveness.

History: Former designation of this paper was SSy-2018-001.

Funding: This material is based on work supported by the National Science Foundation [CAREER Award 1553274].

Keywords: multisecretary problem • regret • adaptive online policy

1. Introduction

In the classic formulation of the secretary problem, a decision maker (referred to as "she") is sequentially presented with n nonnegative, independent values representing the ability of potential candidates and must select one candidate (referred to as "he"). Every time a new candidate is inspected and his ability is revealed, the decision maker must decide whether to reject or select the candidate, and her decision is irrevocable. If the candidate is selected, then the problem ends; if the candidate is rejected, then he cannot be recalled at a later time. The decision maker knows the number of candidates n and the distribution of the ability values in the population, and her objective is to maximize the probability of selecting the most able candidate. For any given n, the problem can be solved by dynamic programming, but there is an asymptotically optimal heuristic that is remarkably elegant. The decision maker observes the abilities of the first n/e candidates and selects the first candidate whose ability exceeds that of the current best candidate, or the last candidate if no such candidate exists (see, e.g., Lindley 1961, Chow et al. 1964, Gilbert and Mosteller 1966, Louchard and Bruss 2016).

Several variations of this simple model have been introduced in the literature, and we refer to Freeman (1983) and Ferguson (1989) for a survey of extensions and references. Relevant to us is the formulation in which the decision maker seeks to maximize the expected ability of the selected candidate rather than maximizing the probability of selecting the best. This problem was first considered by Cayley (1875) and Moser (1956), and it is a special case of the k-choice (multisecretary) problem we study here. In our formulation,

- candidate abilities are independent, identically distributed, and supported on a finite set;
- the decision maker is allowed to select up to k candidates (k is the *recruiting budget*); and
- the decision maker's goal is to maximize the expected total ability of the selected candidates.

This multisecretary problem has applications in revenue management and auctions, among others. In the standard capacity-allocation revenue management problem, a firm sells k items (e.g., airplane seats) to n customers from a discrete set of fare classes over a finite horizon and wishes to allocate the seats in the best possible way (see, e.g., Kleywegt and Papastavrou 1998, Talluri and van Ryzin 2004). In auctions, the decision maker observes arriving bids and must decide whether to allocate one of the available k items to an arriving customer (see, e.g., Kleinberg 2005, Babaioff et al. 2007).

The performance of any online algorithm for this k-choice secretary problem is bounded above by the offline (or posterior) sort: the decision maker waits until all the ability values are presented, sorts them, and then picks the largest k values. As such, we define the "regret" of an online selection algorithm as the expected gap in performance between the online decision and the offline solution, and we prove that the optimal online algorithm has a regret that is bounded uniformly over (n,k). The constant bound depends only on the maximal element of the support and on the minimal mass value of the ability distribution.

Our proof is constructive: we devise an adaptive policy—the budget-ratio (BR) policy—that achieves bounded regret. The policy is adaptive in the sense that the actions are adjusted based on the remaining number of candidates to be inspected and on the residual budget. The proof that the BR policy achieves bounded regret is based on an interesting drift argument for the BR process that tracks the ratio between the residual budget and the remaining number of candidates to be inspected. Under our policy, BR sample paths, as well as their averages, are attracted to and then remain pegged to a certain value (see Section 4). Drift arguments are typical in the study of stability of queues (see, e.g., Bramson et al. 2008), but the proof here is somewhat subtle: the drift is strong early in the horizon, but it weakens as the remaining number of steps decreases. Because "wrong" decisions early in the horizon are more detrimental, this diminishing strength does not compromise the regret. This result shows, in particular, that the BR is a sufficient state descriptor for the development of near-optimal policies. In other words, although there are pairs of (*residual budget*, *remaining number of candidates*) that share the same ratio and the same BR decisions but different optimal actions, these different decisions have little effect on the total expected reward.

We also show that *adaptivity is key*. Although nonadaptive policies could have temporal variation in actions (and hence could have different actions toward the end of the horizon), this variation is not enough: the regret of nonadaptive policies is, in general, of the order of $\sqrt{n}$. Specifically, nonadaptive policies tend to be too greedy and run out of budget too soon. Nonadaptive policies introduce independence between decisions and, consequently, too much variance into the speed at which the recruiting budget is consumed.

Closely related to our work is the paper by Wu et al. (2015), which offers an elegant adaptive-index policy that we revisit in Section 4.1. Under this policy, the ratio of remaining budget to remaining number of steps is a martingale. When the initial ratio of budget to horizon k/n is safely far from the jumps of the discrete distribution, this martingale property guarantees a bounded regret. In general, however, it is precisely the martingale symmetry that increases the regret. In their proof, Wu et al. (2015) show that their policy achieves, up to a constant, a deterministic-relaxation upper bound. This upper bound is not generally achievable, and we must instead use the stochastic offline sort as a benchmark. The detailed analysis of the offline sort gives rise to sufficient conditions that, when satisfied by an online policy, guarantee uniformly bounded regret.

Notation. We use $\mathbb{Z}_+$ to denote the nonnegative integers and $\mathbb{R}_+$ to denote the nonnegative reals. For $j \in \{1,2,\dots\}$, we use $[j]$ to denote the set of integers $\{1,\dots,j\}$, and we set $[j]=\varnothing$ otherwise. Given the real numbers x,y,z, we set $(x)_+ = \max\{0,x\}$, $x \wedge y = \min\{x,y\}$, and we write $y = x \pm z$ to mean that $|y-x| \le z$. Throughout, to simplify notation, we use $M \equiv M(x,y,z)$ to denote a Hardy-style constant dependent on x,y, and z that may change from one line to the next. Finally, we say that a function $\phi(n) = \Theta(\psi(n))$ if there are constants c_1, c_2, and n_0 such that $c_1 \le \frac{\phi(n)}{\psi(n)} \le c_2$ for all $n \ge n_0$.

2. The Multisecretary Problem

A decision maker is sequentially presented with n candidates with abilities $X_1, X_2, \dots, X_n$ and, given a recruiting budget equal to k, can select up to k candidates to maximize the total expected ability of those selected up to and including time n. Of course, there is nothing to study if $k > n$: the decision maker can take all candidates, so it suffices to consider pairs (n,k) that belong to the lattice triangle

$$\mathcal{T} = \{(n,k) \in \mathbb{Z}_+^2 : 0 \le k \le n\}.$$

The abilities $X_1, X_2, \dots, X_n$ are assumed to be independent across candidates and drawn from a common cumulative distribution function F supported on a finite set $\mathcal{A} = \{a_m, a_{m-1}, \dots, a_1 : 0 < a_m < a_{m-1} < \cdots < a_1\}$ of distinct real numbers. We denote by $(f_m, f_{m-1}, \dots, f_1)$ the probability mass function with $f_j = \mathbb{P}(X_1 = a_j)$ for all $j \in [m]$, and

we let F be the cumulative distribution function (so that $f_m = F(a_m) < F(a_{m-1}) < \ldots < F(a_1) = 1$) and $\bar{F} = 1 - F$ be the survival function. Also, for future reference, we choose a value $a_{m+1} < a_m$ with $f_{m+1} = 0$ so that $F(a_{m+1}) = 0$ and $\bar{F}(a_{m+1}) = 1$.

The selection process unfolds as follows: suppose that at time $t \in [n]$, the residual recruiting budget is κ, and the sum of the abilities of the selected candidates up to and including time $t-1$ is w. If the candidate inspected (or "interviewed") at time t has ability $X_t = a$, then the decision maker may *select* the candidate, increasing the cumulative ability to $w + a$ and reducing the residual budget to $\kappa - 1$, or *reject* the candidate, leaving the accrued ability at w and the remaining budget at κ.

A policy is *feasible* if the number of selected candidates does not exceed the recruiting budget k. It is online if the decision with regard to the tth candidate is based only on his ability, the abilities of prior candidates, and the history of the (possibly randomized) decisions up to time t. All decisions are final: if the candidate interviewed at time t is rejected, he is forever lost. Vice versa, if the tth candidate is selected at time t, then that decision cannot be revoked at a later time.

Given a sequence $U_1, U_2, \ldots, U_n$ of independent random variables with the uniform distribution on $[0,1]$ that is also independent of $\{X_1, X_2, \ldots, X_n\}$, we let $\mathcal{F}_0$ denote the trivial σ-field, and for $t \in [n]$, we let $\mathcal{F}_t = \sigma\{X_1, U_1, X_2, U_2, \ldots, X_t, U_t\}$ be the σ-field generated by the random variables $\{X_1, U_1, X_2, U_2, \ldots, X_t, U_t\}$. An online policy π is a sequence of $\{\mathcal{F}_t : t \in [n]\}$-adapted binary random variables $\sigma_1^\pi, \sigma_2^\pi, \ldots, \sigma_n^\pi$, where $\sigma_t^\pi = 1$ means that the candidate with ability X_t is selected. Because the uniform random variables $U_1, U_2, \ldots, U_t$ are included in $\mathcal{F}_t$, the time t decision σ_t^π can be randomized. A feasible online policy requires that the number of selected candidates does not exceed the recruiting budget, that is, that $\sum_{t\in[n]} \sigma_t^\pi \le k$, so

$$\Pi(n,k) \equiv \left\{ (\sigma_1^\pi, \sigma_2^\pi, \ldots, \sigma_n^\pi) \in \{0,1\}^n : \sigma_t^\pi \in \mathcal{F}_t \text{ for all } t \in [n] \text{ and } \sum_{t\in[n]} \sigma_t^\pi \le k \right\}$$

is then the set of all feasible online policies. For $\pi \in \Pi(n,k)$, we let $W_0^\pi, W_1^\pi, \ldots, W_n^\pi$ be the sequence of random variables that track the accumulated ability: we set $W_0^\pi = 0$, and for $r \in [n]$, we let

$$W_r^\pi = \sum_{t\in[r]} X_t \sigma_t^\pi = W_{r-1}^\pi + X_r \sigma_r^\pi.$$

The expected ability accrued at time n by policy π is then given by

$$V_{\text{on}}^\pi(n,k) = \mathbb{E}[W_n^\pi] = \mathbb{E}\left[\sum_{t\in[n]} X_t \sigma_t^\pi \right].$$

For each $(n,k) \in \mathcal{T}$, the goal of the decision maker is to maximize the expected value

$$V_{\text{on}}^*(n,k) = \max_{\pi\in\Pi(n,k)} V_{\text{on}}^\pi(n,k).$$

For completeness, we include in Appendix C an analysis of the dynamic program. The analysis of the Bellman equation confirms an intuitive property of "good" solutions: the optimal action should depend only on the remaining number of steps and the remaining budget and not on the current level of accrued ability. It also allows for a comparison of the BR policy we propose in Section 4 with the optimal policy.

Nonadaptive policies are an interesting subset of feasible online policies. If the residual recruiting budget at time t is positive, then a nonadaptive policy selects the arriving candidate with ability $X_t = a_j$ with probability $p_{j,t} \in [0,1]$ independently of all the previous actions. The probabilities $p_{j,t}$ are determined in advance: they can vary from one period to the next, but this variation is not adapted in response to the previous selection/rejection decisions. A more complete description of nonadaptive policies appears in Section 5. We let $\Pi_{\text{na}} \subseteq \Pi(n,k)$ be the family of nonadaptive policies and define

$$V_{\text{na}}^*(n,k) = \sup_{\pi\in\Pi_{\text{na}}} V_{\text{on}}^\pi(n,k)$$

to be the optimal performance among these.

No feasible online policy—adapted or not—can do better than the offline, full-information counterpart in which all values are presented in advance. The expected ability accrued at time n by the offline problem is given by

$$V_{\text{off}}^*(n,k) = \mathbb{E}\left[\max_{\sigma_1,\ldots,\sigma_n} \left\{ \sum_{t\in[n]} X_t \sigma_t : (\sigma_1, \ldots, \sigma_n) \in \{0,1\}^n \text{ and } \sum_{t\in[n]} \sigma_t \le k \right\} \right].$$

Because the offline solution selects the best k candidates *for each* realization of $X_1, \ldots, X_n$, we have that

$$V^{\pi}_{\text{on}}(n,k) \le V^{*}_{\text{off}}(n,k), \text{ for all } (n,k) \in \mathcal{T} \text{ and all } \pi \in \Pi(n,k).$$

We use the value of the offline problem as a benchmark. The optimal online policy trivially achieves the offline benchmark if $k = 0$ or $k = n$. The main results of this paper (Corollary 1 and Theorem 3) are gathered in the following theorem.

Theorem 1 (The Regret of Online Feasible Policies). *Let* $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$ *. Then there exists a policy* $\text{br} \in \Pi(n,k)$ *and a constant* $M \equiv M(\epsilon)$ *such that*

$$V^{*}_{\text{off}}(n,k) - V^{*}_{\text{on}}(n,k) \le V^{*}_{\text{off}}(n,k) - V^{\text{br}}_{\text{on}}(n,k) \le a_1 M \quad \textit{for all } (n,k) \in \mathcal{T}.$$

Furthermore, if $(f_1 + \epsilon)n \le k \le (1 - f_m - \epsilon)n$*, then an optimal nonadaptive policy has regret that grows like* $\sqrt{n}$ *. Specifically, there is a constant* $M \equiv M(\epsilon, m, a_1, \ldots, a_m)$ *such that*

$$M\sqrt{n} \le V^{*}_{\text{off}}(n,k) - V^{*}_{\text{na}}(n,k).$$

We conclude this section with a discussion of the dependence of the regret on the minimal mass $2\epsilon = \min\{f_m, \ldots, f_1\}$. The following lemma shows that, in general, the regret of the optimal policy grows with ϵ. The example was provided to us by Kleinberg (2017). The proof is in Appendix A.

Lemma 1 (Regret Lower Bound in Kleinberg's (2017) Example). *Consider the three-point distribution on* $\mathcal{A} = \{a_3, a_2, a_1\}$ *with probability mass function* $(\frac{1}{2} + 2\epsilon, 2\epsilon, \frac{1}{2} - 4\epsilon)$*, and let* $n_\epsilon = \lceil \frac{1}{\epsilon^2} \rceil$ *and* $k_\epsilon = \lceil n_\epsilon/2 \rceil$ *. Then there exists a constant* $\Gamma > 0$ *such that*

$$\frac{\Gamma}{\epsilon} \le V^{*}_{\text{off}}(n_\epsilon, k_\epsilon) - V^{*}_{\text{on}}(n_\epsilon, k_\epsilon) \le V^{*}_{\text{off}}(n_\epsilon, k_\epsilon) - V^{\text{br}}_{\text{on}}(n_\epsilon, k_\epsilon).$$

Thus, we also have that

$$\frac{\Gamma}{\epsilon} \le \sup_{(n,k)\in\mathcal{T}} \{V^{*}_{\text{off}}(n,k) - V^{*}_{\text{on}}(n,k)\} \le \sup_{(n,k)\in\mathcal{T}} \{V^{*}_{\text{off}}(n,k) - V^{\text{br}}_{\text{on}}(n,k)\}.$$

Thus, one cannot generally expect a bound that does not depend on the minimal mass. Figure 1 shows that in this case, too; the BR policy we develop in this paper performs remarkably well relative to the optimal. It is just that as ϵ shrinks, the regret of both grows linearly with $1/\epsilon$.

Remark 1 (On the Regret with Uniform Random Permutations). We note here that there are regret upper bounds for a version of this multisecretary problem in which the values are given by a uniform random permutation of the integers $[n]$ instead of being from a random sample. Within the uniform random permutation framework, Kleinberg (2005) proves that the minimal regret in this setting is of the order of $\sqrt{k}$ and provides an algorithm that achieves this lower bound.

3. The Offline Problem

Denoting by $Z^r_j = \sum_{t\in[r]} \mathbb{1}(X_t = a_j)$ the number of candidates with ability a_j inspected up to and including time r, the offline optimization problem has the compact representation

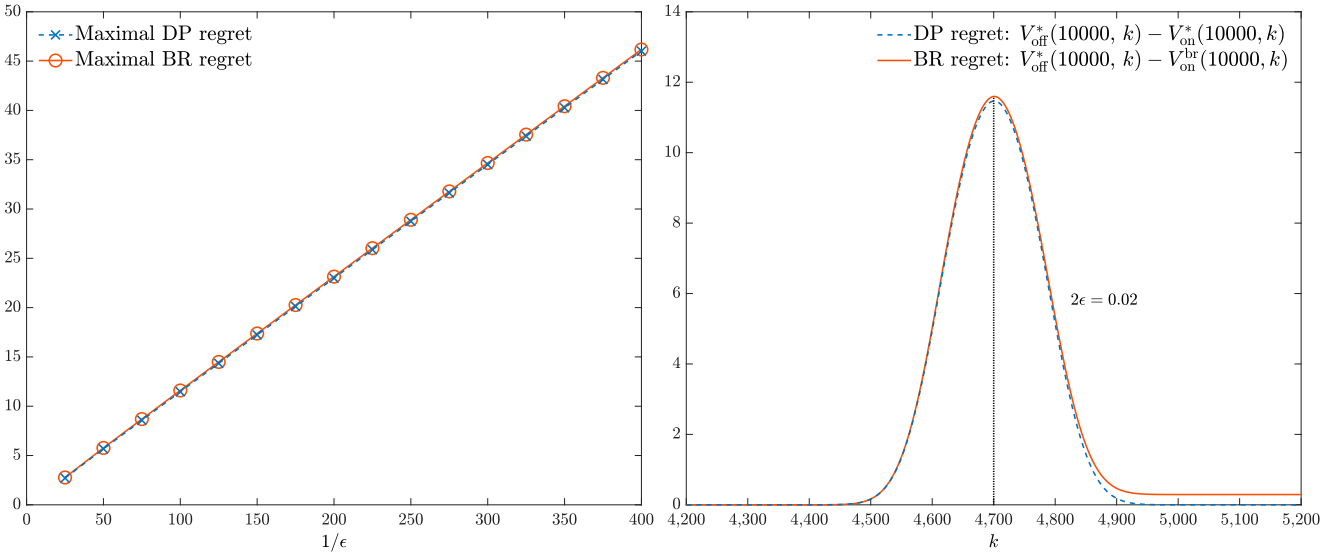
$$V^{*}_{\text{off}}(n,k) = \mathbb{E}[\varphi(Z^n_1, \ldots, Z^n_m, k)], \tag{1}$$

where, for $(z_1, \ldots, z_m) \in \mathbb{Z}^m_+$ and $k \in \mathbb{Z}_+$,

$$\begin{aligned} \varphi(z_1, \ldots, z_m, k) = \max_{s_1,\ldots,s_m} & \sum_{j\in[m]} a_j s_j \\ \text{s.t.} \quad & 0 \le s_j \le z_j \text{ for all } j \in [m], \\ & \sum_{j\in[m]} s_j \le k. \end{aligned} \tag{2}$$

For a given realization of $Z^n_1, \ldots, Z^n_m$, the (trivial) optimal solution is to sort the values and select the k candidates with the largest abilities. First, one selects $\mathfrak{S}^n_1 = \min\{Z^n_1, k\}$ candidates with ability a_1. Then, if there is

Figure 1. Dynamic-Programming and BR Regrets in Kleinberg's (2017) Example



Notes. The figure shows Kleinberg's (2017) example on $\mathcal{A} = \{1, 2, 3\}$ and probability mass function $(\frac{1}{2} + 2\epsilon, 2\epsilon, \frac{1}{2} - 4\epsilon)$. On the left, we see that the regret grows linearly with $1/\epsilon$ (as $n_\epsilon = \lceil 1/\epsilon^2 \rceil$). On the right, we take $\epsilon = 0.01$ and $n = 10,000$ and note that the regret of both the dynamic-programming (DP) and the BR policies is maximized at $k = 4{,}700$, which corresponds to the threshold $T_2 = \frac{1}{2}(\bar{F}(a_2) + \bar{F}(a_3)) = 0.47$, which will be key in our BR policy for this setting.

any recruiting budget left, one selects candidates with ability a_2 until either all of them are selected or the remaining budget of $k - Z_1^n$ is exhausted; that is, one selects $\mathfrak{S}_2^n = \min\{Z_2^n, (k - Z_1^n)_+\}$ candidates with ability a_2. In general, the offline number of selected a_j candidates is given by

$$\mathfrak{S}_j^n = \min\left\{Z_j^n, \left(k - \sum_{i \in [j-1]} Z_i^n\right)_+\right\} \qquad \text{for each } j \in [m], \tag{3}$$

so

$$V^*_{\text{off}}(n,k) = \sum_{j\in[m]} a_j \mathbb{E}[\,\mathfrak{S}_j^n] = \sum_{j\in[m]} a_j \mathbb{E}\left[\min\left\{Z_j^n, \left(k - \sum_{i \in [j-1]} Z_i^n\right)_+\right\}\right]. \tag{4}$$

The offline solution (3) has the appealing property that up to deviations that are constant in expectation, all the action is within two ability levels; that is, depending on the ratio k/n, there is an index $j_0 \in [m]$ such that the offline sort algorithm rejects almost all the candidates with ability strictly below a_{j_0+1} and selects all the candidates with ability strictly above a_{j_0}. The action index j_0 depends on the horizon length n and the recruiting budget k, and it is given for any pair $(n,k) \in \mathcal{T}$ by

$$j_0(n,k) = \begin{cases} 1 & \text{if } k/n < f_1 + f_2/2, \\ j : \bar{F}(a_j) + f_j/2 \le k/n < \bar{F}(a_{j+2}) - f_{j+1}/2 & \text{if } f_1 + f_2/2 \le k/n < 1 - f_m/2, \\ m & \text{if } 1 - f_m/2 \le k/n. \end{cases} \tag{5}$$

If (n,k) are such that $j_0(n,k) = j$, then the two values at play are a_j and a_{j+1}, and this suggests partitioning $\mathcal{T}$ into the sets

$$\mathcal{T}_j = \{(n,k) \in \mathcal{T} : j_0(n,k) = j\}, \quad j \in \{1, 2, \ldots, m\}$$

to obtain a helpful decomposition of the offline value (4).

Proposition 1 (Offline Sort Decomposition). *Let $\epsilon = \frac{1}{2}\min\{f_m, \ldots, f_1\}$, and fix $j \in \{1, \ldots, m\}$. For all $(n,k) \in \mathcal{T}_j$, one has the bounds*

$$\sum_{i\in[j-1]} \mathbb{E}[Z_i^n] - \frac{1}{4\epsilon} \le \sum_{i\in[j-1]} \mathbb{E}[\mathfrak{S}_i^n] \qquad \text{and} \qquad \sum_{i=j+2}^{m} \mathbb{E}[\mathfrak{S}_i^n] \le \frac{1}{4\epsilon}. \tag{6}$$

Consequently, for all $(n,k)\in\mathcal{T}_j$, *one has the decomposition*

$$V^*_{\mathrm{off}}(n,k)=\sum_{i\in[j-1]}a_i\mathbb{E}[Z^n_i]+a_j\mathbb{E}[\mathfrak{S}^n_j]+a_{j+1}\mathbb{E}[\mathfrak{S}^n_{j+1}]\pm\frac{a_1}{4\epsilon}. \tag{7}$$

The following lemma will be used repeatedly in what follows. All lemmas stated in this paper are proved in Appendix B.

Lemma 2. *Let* B *be a binomial random variable with* n *trials and success probability* p. *Then, for any* $\varepsilon>0$,

$$\mathbb{E}[(B-k)_+]\le\frac{1}{4\varepsilon}\quad\text{if } p+\varepsilon\le\frac{k}{n},\qquad\text{and}\qquad\mathbb{E}[(k-B)_+]\le\frac{1}{4\varepsilon}\quad\text{if } \frac{k}{n}\le p-\varepsilon. \tag{8}$$

The first use of Lemma 2 is in the proof of Proposition 1.

Proof of Proposition 1. If $j=1$, the left inequality of (6) reduces to $-(4\epsilon)^{-1}\le 0$, and there is nothing to prove. For $1<j$ and $\iota<j$, we obtain from (3) that

$$0\le Z^n_\iota-\mathfrak{S}^n_\iota=\left(\min\left\{Z^n_\iota,\sum_{i\in[\iota]}Z^n_i-k\right\}\right)_+\le\left(\sum_{i\in[\iota]}Z^n_i-k\right)_+\le\left(\sum_{i\in[j-1]}Z^n_i-k\right)_+.$$

By the definition of $\mathcal{T}_j$ and (5), we have that $\bar{F}(a_j)+\epsilon\le k/n$, so, because the sum $B=\sum_{i\in[j-1]}Z^n_i$ is binomial with parameters n and $\bar{F}(a_j)=f_{j-1}+\ldots+f_1$, the first inequality in (6) follows from (8).

For the second inequality in (6), notice that if $j\in\{m-1,m\}$, then there is nothing to prove. Otherwise, if $j\le m-2$, we have, for all $\iota\ge j+2$ (or, equivalently, for $\iota-1\ge j+1$), that

$$\mathfrak{S}^n_\iota=\min\left\{Z^n_\iota,\left(k-\sum_{i\in[\iota-1]}Z^n_i\right)_+\right\}\le\left(k-\sum_{i\in[\iota-1]}Z^n_i\right)_+\le\left(k-\sum_{i\in[j+1]}Z^n_i\right)_+.$$

The sum $B=\sum_{i\in[j+1]}Z^n_i$ is a binomial random variable with success probability $\bar{F}(a_{j+2})=f_{j+1}+f_j+\ldots+f_1$, so, because $\epsilon=\frac{1}{2}\min\{f_m,f_{m-1},\ldots,f_1\}$ and $(n,k)\in\mathcal{T}_j$, we have that $k/n\le\bar{F}(a_{j+2})-\epsilon$, and the second inequality in (6) again follows from (8).

To conclude the proof of the proposition, we recall (4) and rewrite $V^*_{\mathrm{off}}(n,k)$ as

$$V^*_{\mathrm{off}}(n,k)=-\sum_{i\in[j-1]}a_i\{\mathbb{E}[Z^n_i]-\mathbb{E}[\mathfrak{S}^n_i]\}+\sum_{i\in[j-1]}a_i\mathbb{E}[Z^n_i]+a_j\mathbb{E}[\mathfrak{S}^n_j]+a_{j+1}\mathbb{E}[\mathfrak{S}^n_{j+1}]+\sum_{i=j+2}^{m}a_i[\mathfrak{S}^n_i]. \tag{9}$$

The definition of $\mathfrak{S}^n_i$ in (3), the monotonicity $a_m<a_{m-1}<\cdots<a_1$, and the left inequality of (6) then give us that

$$0\le\sum_{i\in[j-1]}a_i\{\mathbb{E}[Z^n_i]-\mathbb{E}[\mathfrak{S}^n_i]\}\le a_1\left(\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]-\sum_{i\in[j-1]}\mathbb{E}[\mathfrak{S}^n_i]\right)\le\frac{a_1}{4\epsilon}. \tag{10}$$

Similarly, the monotonicity $a_m<a_{m-1}<\cdots<a_1$ and the right inequality of (6) then imply that

$$0\le\sum_{i=j+2}^{m}a_i[\mathfrak{S}^n_i]\le\frac{a_1}{4\epsilon}, \tag{11}$$

so the decomposition (7) follows after one estimates the first sum and the last sum of (9) with the bounds in (10) and (11), respectively. □

For any feasible online policy $\pi\in\Pi(n,k)$, we let $S^{\pi,r}_j=\sum_{t=1}^{r}\sigma^\pi_t\mathbb{1}(X_t=a_j)$ be the number of candidates with ability level a_j that are selected by policy π up to and including time r, so the expected total ability accrued by policy π can be written as

$$V^\pi_{\mathrm{on}}(n,k)=\mathbb{E}\left[\sum_{t\in[n]}X_t\sigma^\pi_t\right]=\sum_{j\in[m]}a_j\mathbb{E}[S^{\pi,n}_j].$$

Proposition 1 suggests that to have bounded regret, an online algorithm must be selecting almost all candidates with abilities $a_1, a_2, \ldots, a_{j_0-1}$ and rejecting almost all candidates with abilities $a_{j_0+2}, \ldots, a_m$ if j_0 is such that $k/n \in [\bar{F}(a_{j_0}) + \frac{1}{2}f_{j_0}, \bar{F}(a_{j_0+2}) - \frac{1}{2}f_{j_0+1})$. This sufficient condition guides us in the development of the budget-ratio policy in Section 4.

Proposition 2 (A Sufficient Condition). *Let $\epsilon = \frac{1}{2}\min\{f_m, \ldots, f_1\}$. For any given $(n,k) \in \mathcal{T}$, let $j_0 \equiv j_0(n,k)$ be the index defined by (5), and suppose that there is a feasible online policy $\pi \equiv \pi(n,k) \in \Pi(n,k)$, a stopping time $\tau \equiv \tau(\pi) \le n$, and a constant $M \equiv M(\epsilon) < \infty$ such that*

(i) $\sum_{j\in[j_0-1]} \mathbb{E}[S_j^{\pi,\tau}] = \sum_{j\in[j_0-1]} \mathbb{E}[Z_j^{\tau}]$,

(ii) $\mathbb{E}[S_{j_0}^{\pi,\tau}] \ge \mathbb{E}[\mathfrak{S}_{j_0}^{\tau}] - M$,

(iii) $\mathbb{E}[S_{j_0+1}^{\pi,\tau}] \ge \mathbb{E}[\mathfrak{S}_{j_0+1}^{\tau}] - M$, *and*

(iv) $\mathbb{E}[\tau] \ge n - M$. *Then one has that*

$$\sup_{(n,k)\in\mathcal{T}} \{V_{\text{off}}^*(n,k) - V_{\text{on}}^{\pi}(n,k)\} \le 3a_1M + \frac{a_1}{4\epsilon}.$$

The regret bound in Proposition 2 has two components (summands). The first one accounts for three inefficiencies of the online policy π with respect to the offline solution. There is the cost for running out of budget too early—at most M periods too early in expectation according to condition (iv)—which cannot exceed a_1M, and there are the costs that come from conditions (ii) and (iii) that can each contribute to a maximal expected loss of a_1M. Finally, the second summand accounts for the values that the offline solution might be choosing and that are strictly smaller than a_{j_0+1}. However, Proposition 1 tells us that their cumulative expected value is at most $a_1/(4\epsilon)$.

Proof of Proposition 2. Because $\mathcal{T} = \bigcup_{j\in[m]} \mathcal{T}_j$, we have that

$$\sup_{(n,k)\in\mathcal{T}} \{V_{\text{off}}^*(n,k) - V_{\text{on}}^{\pi}(n,k)\} = \max_{j\in[m]} \left\{ \sup_{(n,k)\in\mathcal{T}_j} \{V_{\text{off}}^*(n,k) - V_{\text{on}}^{\pi}(n,k)\} \right\},$$

and it suffices to verify that for any $j \in [m]$,

$$\sup_{(n,k)\in\mathcal{T}_j} \{V_{\text{off}}^*(n,k) - V_{\text{on}}^{\pi}(n,k)\} \le 3a_1M + \frac{a_1}{4\epsilon}.$$

For any $(n,k) \in \mathcal{T}_j$, the definition (5) of the map $(n,k) \mapsto j_0(n,k)$ gives us that $j_0(n,k) = j$, so for any policy $\pi \in \Pi(n,k)$ and any stopping time $\tau \le n$, we have the lower bound

$$\sum_{i\in[j-1]} a_i\mathbb{E}[S_i^{\pi,\tau}] + a_j\mathbb{E}[S_j^{\pi,\tau}] + a_{j+1}\mathbb{E}[S_{j+1}^{\pi,\tau}] \le \sum_{j\in[m]} a_j\mathbb{E}[S_j^{\pi,n}] = V_{\text{on}}^{\pi}(n,k).$$

In turn, for the policy $\pi \equiv \pi(n,k)$ and the stopping time $\tau \equiv \tau(\pi)$ given in the proposition, it follows from the requirements (i), (ii), and (iii) that there is a constant $M \equiv M(\epsilon)$ such that

$$\sum_{i\in[j-1]} a_i\mathbb{E}[Z_i^{\tau}] + a_j\{\mathbb{E}[\mathfrak{S}_j^{\tau}] - M\} + a_{j+1}\{\mathbb{E}[\mathfrak{S}_{j+1}^{\tau}] - M\} \le V_{\text{on}}^{\pi}(n,k),$$

which, because $a_m < a_{m-1} < \ldots < a_1$, gives

$$\sum_{i\in[j-1]} a_i\mathbb{E}[Z_i^{\tau}] + a_j\mathbb{E}[\mathfrak{S}_j^{\tau}] + a_{j+1}\mathbb{E}[\mathfrak{S}_j^{\tau}] - 2a_1M \le V_{\text{on}}^{\pi}(n,k). \tag{12}$$

For $\tau \le n$, we now rewrite the upper bound in (7) as

$$\begin{aligned} V_{\text{off}}^*(n,k) \le{}& \sum_{i\in[j-1]} a_i\mathbb{E}[Z_i^{\tau}] + a_j\mathbb{E}[\mathfrak{S}_j^{\tau}] + a_{j+1}\mathbb{E}[\mathfrak{S}_{j+1}^{\tau}] + \frac{a_1}{4\epsilon} \\ &+ \sum_{i\in[j-1]} a_i\mathbb{E}[Z_i^{n} - Z_i^{\tau}] + a_j\mathbb{E}[\mathfrak{S}_j^{n} - \mathfrak{S}_j^{\tau}] + a_{j+1}\mathbb{E}[\mathfrak{S}_{j+1}^{n} - \mathfrak{S}_{j+1}^{\tau}], \end{aligned} \tag{13}$$

and we obtain from the definition (3) of the offline number of selected a_i candidates that the difference $0 \le \mathfrak{S}_i^n - \mathfrak{S}_i^\tau \le Z_i^n - Z_i^\tau = \sum_{t=\tau+1}^n \mathbb{1}(X_t = a_i)$ for all $i \in [m]$. Thus, if we recall that $a_m < a_{m-1} < \ldots < a_1$, we obtain the bound

$$\sum_{i\in[j-1]} a_i \mathbb{E}[Z_i^n - Z_i^\tau] + a_j \mathbb{E}[\mathfrak{S}_j^n - \mathfrak{S}_j^\tau] + a_{j+1}\mathbb{E}[\mathfrak{S}_{j+1}^n - \mathfrak{S}_{j+1}^\tau] \le a_1 \mathbb{E}\left[\sum_{i\in[m]} \sum_{t=\tau+1}^n \mathbb{1}(X_t = a_i)\right] = a_1 \mathbb{E}[n-\tau],$$

so when we use property (iv) to estimate this last right-hand side and replace this estimate in (13), we finally have that

$$V_{\text{off}}^*(n,k) \le \sum_{i\in[j-1]} a_i \mathbb{E}[Z_i^\tau] + a_j \mathbb{E}[\mathfrak{S}_j^\tau] + a_{j+1}\mathbb{E}[\mathfrak{S}_{j+1}^\tau] + a_1 M + \frac{a_1}{4\epsilon}. \tag{14}$$

The proof of the proposition then follows by combining the bounds (12) and (14). □

Remark 2 (A Deterministic Relaxation). A further upper bound to the multisecretary problem is provided by an intuitive deterministic relaxation. Given the linear program (2), we consider its optimal value

$$DR(n,k) = \varphi(\mathbb{E}[Z_1^n], \ldots, \mathbb{E}[Z_m^n], k), \tag{15}$$

and we note that its optimal solution is given by

$$s_j^* = \min\left\{\mathbb{E}[Z_j^n], \left(k - \sum_{i\in[j-1]} \mathbb{E}[Z_i^n]\right)_+\right\} = \min\{n f_j, (k - n\bar{F}(a_j))_+\} \qquad \text{for all } j \in [m]. \tag{16}$$

Because the expected number of a_j candidates selected by the offline solution $\widehat{s}_j = \mathbb{E}[\mathfrak{S}_j^n] = \mathbb{E}[\min\{Z_j^n, (k - \sum_{i\in[j-1]} Z_i^n)_+\}]$ is a feasible solution for the deterministic relaxation (15), we immediately have the bound

$$V_{\text{off}}^*(n,k) \le DR(n,k) \quad \text{for all pairs } (n,k) \in \mathcal{T}.$$

As central-limit-theorem intuition suggests, the cost of randomness is at most of the order of $\sqrt{n}$. But it does not have to be that large: there is a subset $\mathcal{T}' \subset \mathcal{T}$ for which the difference $DR(n,k) - V_{\text{off}}^*(n,k)$ is bounded by a constant that does not depend on $(n,k) \in \mathcal{T}'$. Members (n,k) of $\mathcal{T}'$ are such that k/n is "safely" away from the jump points of the distribution F (see Appendix D). For $(n,k) \in \mathcal{T}'$, benchmarking against the deterministic relaxation is the same as benchmarking against the offline sort (see also Wu et al. 2015). In general, this is not the case. For instance, if one takes $k = \mathbb{E}[Z_1^n] = nf_1$, then $DR(n,k) = a_1 n f_1$, but there exists $M < \infty$ such that $a_1 \mathbb{E}[\mathfrak{S}_1^n] = a_1 \mathbb{E}[\min\{Z_1^n, k\}] = a_1 n f_1 + \mathbb{E}[\min\{Z_1^n - nf_1, 0\}] \le a_1 n f_1 - M\sqrt{n} = DR(n,k) - M\sqrt{n}$. The first inequality follows from the approximation of the centered binomial $Z_1^n - nf_1$ by a normal with standard deviation $\sqrt{nf_1(1-f_1)}$, as in Lemma 5.

4. The BR Policy

We now introduce an adaptive online policy that makes selection decisions depending on the ratio between the remaining number of positions to be filled (the remaining budget) and the remaining number of candidates to be inspected (the remaining time), and we refer to this policy as the "budget-ratio (BR) policy."

With $\pi = \text{br}$, the random variables $\sigma_1^{\text{br}}, \sigma_2^{\text{br}}, \ldots, \sigma_n^{\text{br}}$ give us the sequence of selection decisions under the BR policy (see Section 2), and we let, for $t \in [n]$,

$$K_t = k - \sum_{s\in[t]} \sigma_s^{\text{br}} = K_{t-1} - \sigma_t^{\text{br}}$$

be the remaining budget after the tth decision ($K_0 = k$). We now introduce the thresholds $0 = T_1 < T_2 < \cdots < T_m < T_{m+1} = +\infty$ given by

$$T_j = \frac{1}{2}(\bar{F}(a_j) + \bar{F}(a_{j+1})) \qquad \text{for each } j \in \{2, 3, \ldots, m\},$$

so the BR decision at time $t+1$ selects X_{t+1} depending on its value and on the position of the ratio $K_t/(n-t)$ relative to these thresholds. Specifically, at each decision time $t+1 \in \{1, \ldots n-1\}$, the BR policy

1. identifies the index $j \in [m]$ such that

$$T_j \le \frac{K_t}{n-t} < T_{j+1};$$

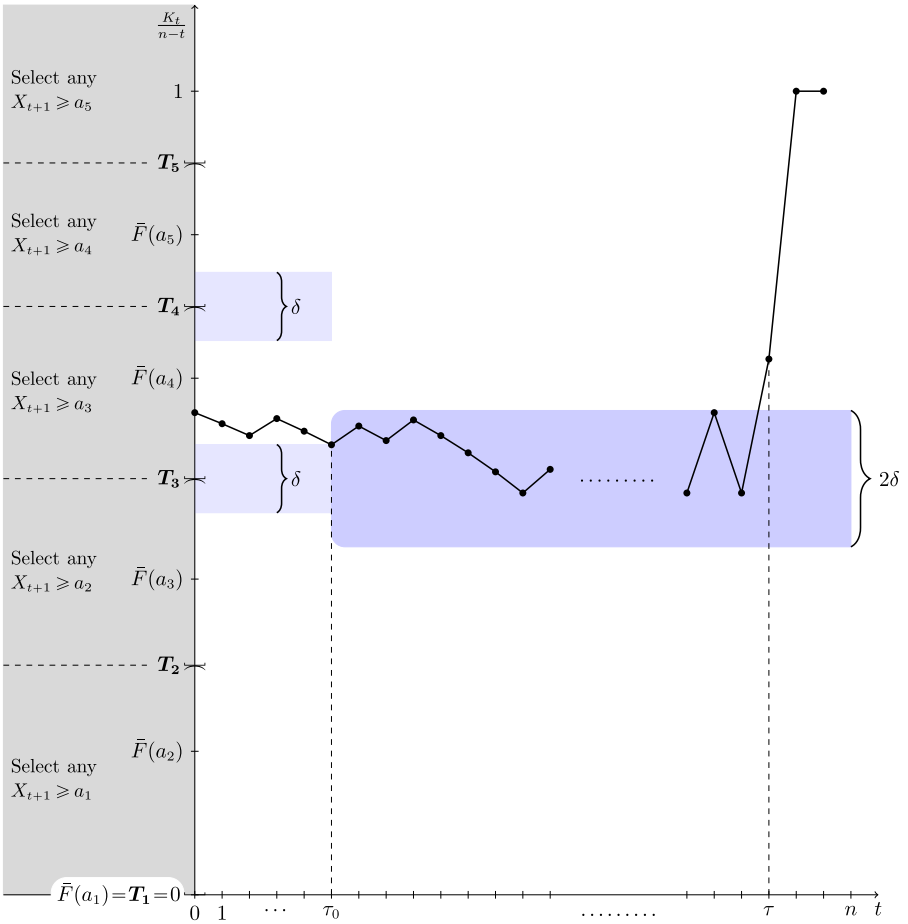
2. selects X_{t+1} if and only if $K_t > 0$ and $X_{t+1} \ge a_j$, that is, sets

$$\sigma_{t+1}^{\mathrm{br}} = \begin{cases} 1 & \text{if } K_t > 0 \text{ and } X_{t+1} > a_{j+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the value X_{t+1} is selected with probability $\mathbb{P}(\sigma_{t+1}^{\mathrm{br}} = 1 \,|\, \mathcal{F}_t) = \bar{F}(a_{j+1})\mathbb{1}(K_t > 0)$.

Figure 2 gives a graphical representation of the selection regions of the BR policy. It encapsulates the source of its good performance. Consider an initial budget ratio $k/n \in [T_3, \bar{F}(a_4))$ as in the figure. In this range, there is enough budget, in expectation, to select all the candidates with ability values a_1 and a_2. The remaining budget after such selections should be used for candidates with ability equal to a_3 or smaller. While the budget ratio is in the colored band, the policy will select all the a_1 and a_2 values. It will also select a_3 candidates whenever the budget ration is above T_3. If we can guarantee that the budget ratio stays in the colored band almost until the end of the horizon, then the BR policy will select as many a_3 values as possible without giving up on any a_1 or a_2 values. This logic is formalized in the proof of Corollary 1. The policy has two natural properties: (i) because

Figure 2. The BR Policy: Thresholds and Dynamics



Notes. The y-axis shows the thresholds of the BR policy for the five-point distribution on $\mathcal{A} = \{a_5, a_4, a_3, a_2, a_1\}$ with the probability mass function $(f_5, f_4, f_3, f_2, f_1) = (\frac{5}{28}, \frac{5}{28}, \frac{7}{28}, \frac{6}{28}, \frac{5}{28})$. The plotted series is a sample-path realization of the ratio $\{\frac{K_t}{n-t} : 0 \le t \le n\}$, which enters the orbit of the threshold T_3 at time τ_0 (so $j(\tau_0) = 3$) and exits at time τ. Until time τ_0, both thresholds T_3 and T_4 are in play. In this chart, we take $\delta = \frac{19}{224} < \epsilon = \frac{5}{56} = \frac{1}{2}\min\{f_5, \ldots, f_1\}$. Notice that $\bar{F}(a_3) = f_1 + f_2$. When $\bar{F}(a_3) < K_t/(n-t) < T_3$, the budget is, in expectation, sufficient to take some a_3 values, but the policy will not do that until T_3 is crossed. This "underselection" makes $K_t/(n-t)$ drift up toward T_3. When $T_3 < K_t/(n-t) < \bar{F}(a_4)$, the budget is, in expectation, insufficient to take all a_3 values, but the policy does select them. This "overselection" makes $K_t/(n-t)$ drift toward T_3.

$T_1 = 0$, the BR policy selects all a_1-valued candidates until exhausting the budget, and (ii) because $T_{m-1} < 1$ and $T_m = +\infty$, the BR policy selects all remaining values as soon as the remaining budget is greater than or equal to the remaining number of time periods (i.e., if $n - t \le K_t$).

Next, we set

$$\epsilon = \tfrac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\},$$

fix $0 < \delta < \epsilon$, and consider the stopping time

$$\tau_0 = \inf\left\{t \ge 0 : \left|\frac{K_t}{n-t} - T_j\right| \le \frac{\delta}{2} \text{ for some } j \in [m] \text{ or } t \ge n - 2\delta^{-1} - 1\right\}.$$

If $\tau_0 < n - 2\delta^{-1} - 1$, then τ_0 is the first time that the ratio $K_t/(n-t)$ enters the "orbit" of one of the thresholds (see Figure 2). We denote by $j(\tau_0)$ the index of the threshold that is within $\delta/2$ of the ratio $K_{\tau_0}/(n-\tau_0)$, and we use $T_{j(\tau_0)}$ to denote the value of that threshold. If $\tau_0 = n - 2\delta^{-1} - 1$, then we set $j(\tau_0) = m + 1$ and $T_{j(\tau_0)} = \infty$. For all $t \le n - 2\delta^{-1} - 1$, the jumps of $K_t/(n-t)$ satisfy the absolute bound

$$\left|\frac{K_t}{n-t} - \frac{K_{t+1}}{n-(t+1)}\right| \le \frac{\delta}{2},$$

so that, in the event $\tau_0 < n - 2\delta^{-1} - 1$, we are guaranteed that $T_{j(\tau_0)}$ is either T_j or T_{j+1} when j is such that $k/n \in [T_j, T_{j+1})$.

After time τ_0, we consider the process

$$Y_u = K_{\tau_0+u} - T_{j(\tau_0)}(n - \tau_0 - u) \qquad \text{for } u \in \{0, 1, \ldots n - \tau_0\},$$

which serves as a useful vehicle to study the behavior of the BR process and, in particular, to track the deviations of the BR from the threshold $T_{j(\tau_0)}$. This is because

$$\left|\frac{K_{\tau_0+u}}{n-\tau_0-u} - T_{j(\tau_0)}\right| > \delta \quad \text{if and only if} \quad |Y_u| > \delta(n - \tau_0 - u).$$

In words, the ratio is outside the dark region in Figure 2 if and only if the deviation process Y_u exceeds the "moving target" $\delta(n - \tau_0 - u)$.

The process $\{Y_u : 0 \le u \le n - \tau_0\}$ is adapted to the increasing sequence of σ-fields $\{\widehat{\mathcal{F}}_u \equiv \mathcal{F}_{\tau_0+u} : 0 \le u \le n - \tau_0\}$. Importantly, the process $\{Y_u : 0 \le u \le n - \tau_0\}$ has the mean-reversal property that we alluded to in the description of Figure 2: if $K_{\tau_0+u} > 0$ and $Y_u \ge 0$, then we have that $T_{j(\tau_0)} \le K_{\tau_0+u}/(n - \tau_0 - u)$, and the BR policy selects *all* values strictly larger than $a_{j(\tau_0)+1}$, that is, $\sigma^{\mathrm{br}}_{\tau_0+u+1} \ge \mathbb{1}(X_{\tau_0+u+1} > a_{j(\tau_0)+1})$ so that

$$\mathbb{E}\left[\sigma^{\mathrm{br}}_{\tau_0+u+1} \,|\widehat{\mathcal{F}}_u\right]\mathbb{1}(K_{\tau_0+u} > 0, Y_u \ge 0) \ge \bar{F}(a_{j(\tau_0)+1})\mathbb{1}(K_{\tau_0+u} > 0, Y_u \ge 0),$$

and we have the negative-drift property

$$\begin{aligned}\mathbb{E}[Y_{u+1} - Y_u \,|\widehat{\mathcal{F}}_u]\mathbb{1}(K_{\tau_0+u} > 0, Y_u \ge 0) &= \left\{-\mathbb{E}[\sigma^{\mathrm{br}}_{\tau_0+u+1} \,|\widehat{\mathcal{F}}_u] + T_{j(\tau_0)}\right\}\mathbb{1}(K_{\tau_0+u} > 0, Y_u \ge 0) \\ &\le -\frac{1}{2} f_{j(\tau_0)}\mathbb{1}(K_{\tau_0+u} > 0, Y_u \ge 0).\end{aligned} \tag{17}$$

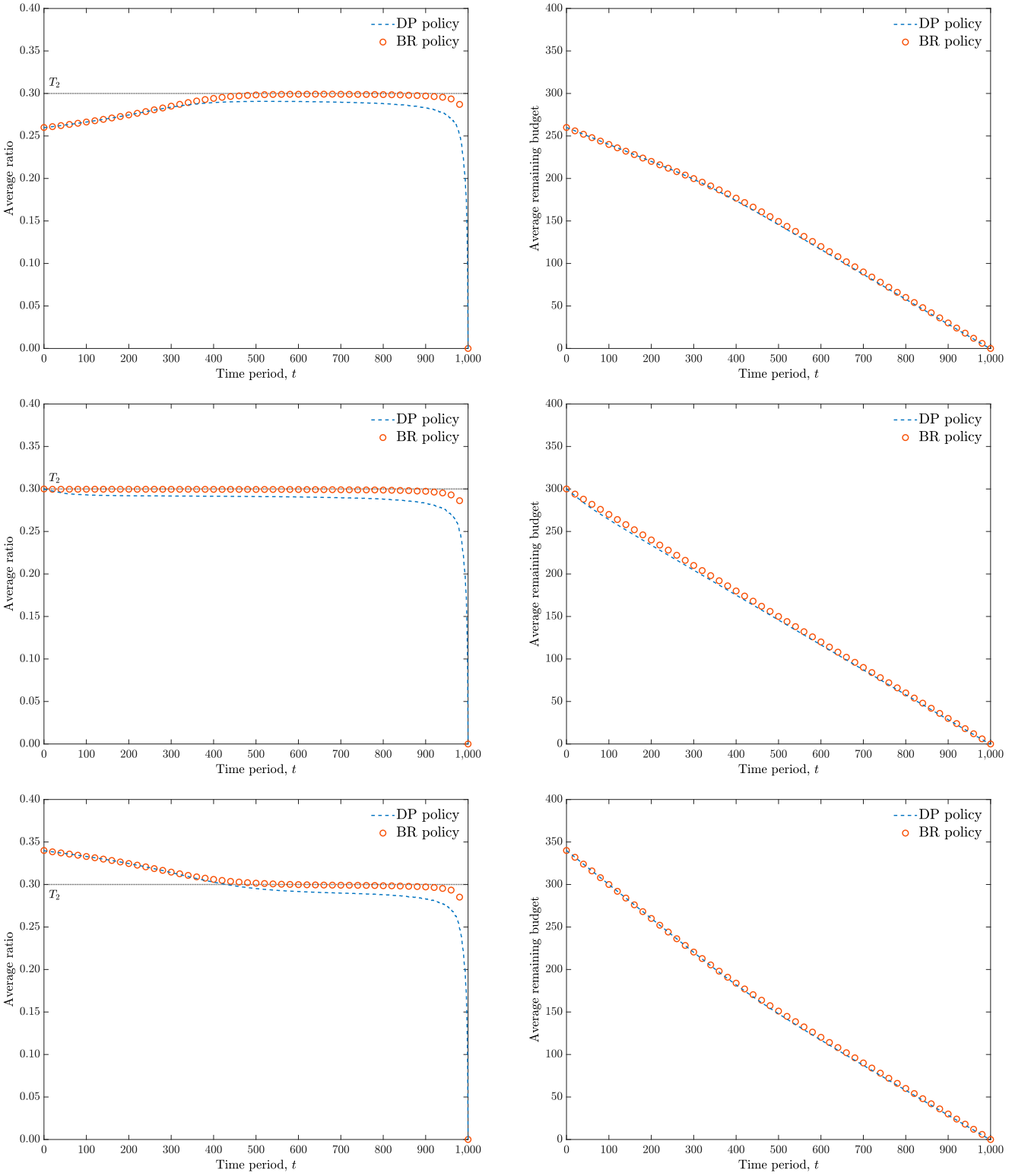
In the case that $Y_u < 0$ and $K_{\tau_0+u} > 0$, we have that $K_{\tau_0+u}/(n - \tau_0 - u) < T_{j(\tau_0)}$, so the BR policy skips all values smaller than or equal to $a_{j(\tau_0)}$; that is, $\sigma^{\mathrm{br}}_{\tau_0+u+1} \le \mathbb{1}(X_{\tau_0+u+1} > a_{j(\tau_0)})$ so that

$$\mathbb{E}[\sigma^{\mathrm{br}}_{\tau_0+u+1} \,|\widehat{\mathcal{F}}_u]\mathbb{1}(K_{\tau_0+u} > 0, Y_u < 0) \le \bar{F}(a_{j(\tau_0)})\mathbb{1}(K_{\tau_0+u} > 0, Y_u < 0),$$

and we have a strictly positive drift

$$\begin{aligned}\mathbb{E}[Y_{u+1} - Y_u \,|\widehat{\mathcal{F}}_u]\mathbb{1}(K_{\tau_0+u} > 0, Y_u < 0) &= \left\{-\mathbb{E}[\sigma^{\mathrm{br}}_{\tau_0+u+1} \,|\widehat{\mathcal{F}}_u] + T_{j(\tau_0)}\right\}\mathbb{1}(K_{\tau_0+u} > 0, Y_u < 0) \\ &\ge \frac{1}{2} f_{j(\tau_0)}\mathbb{1}(K_{\tau_0+u} > 0, Y_u < 0).\end{aligned} \tag{18}$$

Figure 3. Simulated Ratio and Remaining Budget Averages



Notes. The figure shows simulated averages—with $n = 1{,}000$ and based on 10,000 trials—of the ratio $\frac{1}{n-t}(k - \sum_{s\in[t]}\sigma_s^\pi)$ and of the remaining budget $k - \sum_{s\in[t]}\sigma_s^\pi$ when π is the optimal dynamic-programming (DP) policy or the BR policy. The ability distribution is uniform on $\mathcal{A} = \{0.20, 0.65, 1.10, 1.55, 2.00\}$. In the plots, we vary the budget $k \in \{260, 300, 340\}$ starting with the smallest at the top and while keeping the ratio k/n within the orbit of $T_2 = 0.30$.

For completeness, we also note that if $K_{\tau_0+u} = 0$, then the drift is simply given by

$$\mathbb{E}[\,Y_{u+1} - Y_u \,|\widehat{\mathcal{F}}_u]\mathbb{1}(K_{\tau_0+u} = 0) = T_{j(\tau_0)}\mathbb{1}(K_{\tau_0+u} = 0). \tag{19}$$

The bounds (17) and (18) show that regardless of whether the ratio $K_{\tau_0+u}/(n-\tau_0-u)$ is above or below the critical threshold, the BR policy pulls it toward the threshold. This preliminary drift analysis will be useful in showing that the stopping time

$$\tau = \inf\left\{t > \tau_0 : \left|\frac{K_t}{n-t} - T_{j(\tau_0)}\right| > \delta \;\text{ or }\; t \geq n - 2\delta^{-1} - 1\right\}, \tag{20}$$

at which the ratio exits the critical orbit (see again the right side of Figure 2), is suitably large.

Theorem 2 (BR Stopping Time). *Let $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$. Then there is a constant $M \equiv M(\epsilon)$ such that for all $(n,k) \in \mathcal{T}$, the stopping time τ in (20) satisfies the bound $\mathbb{E}[\tau] \geq n - M$.*

The proof of Theorem 2 (at the end of this section) is based on the mean-reversal property established earlier and a Lyapunov function argument. One must be careful in this analysis: when approaching the horizon's end, a small change in K_t can lead to a large change in $K_t/(n-t)$. Viewed in terms of the deviation process Y_u, the challenge is that the target $\delta(n-\tau_0-u)$ is moving and easier to exceed when u is large. Figure 4 is a simulation that illustrates how this "attraction" to a threshold is manifested at the sample-path level.

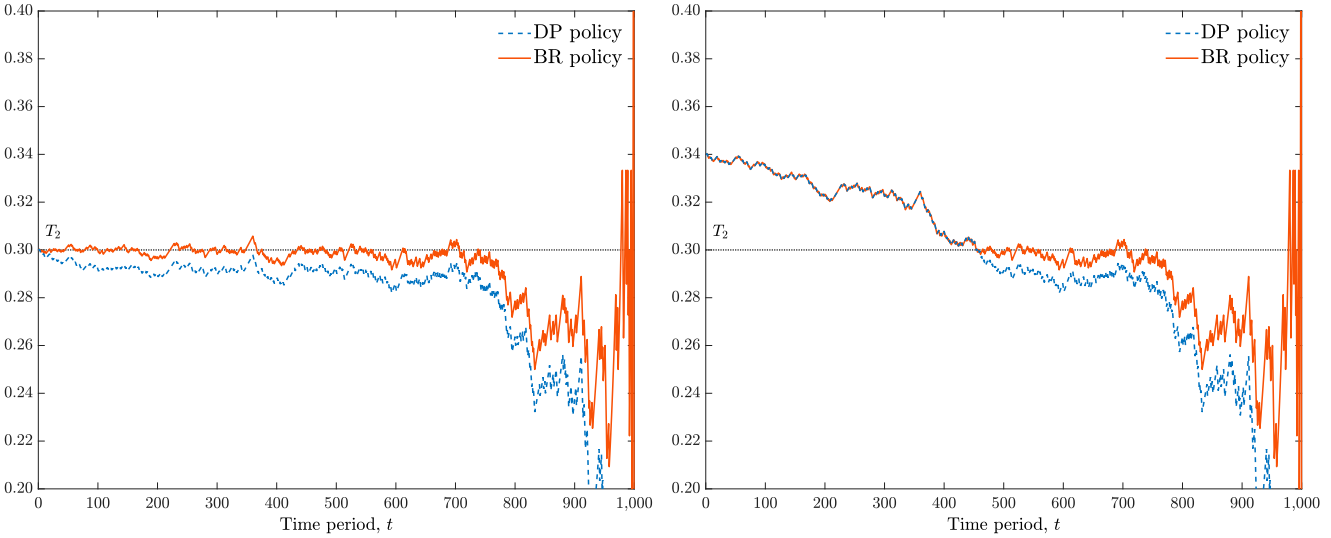
Theorem 2 shows that the BR policy satisfies the sufficient condition (iv) in Proposition 2. Corollary 1 then proceeds to show that the remaining requirements (i)–(iii) in Proposition 2 are also satisfied.

Corollary 1 (Uniformly Bounded Regret). *Let $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$. Then the BR policy and the stopping time τ in (20) satisfy the properties in Proposition 2. In particular, there is a constant $M \equiv M(\epsilon)$ such that*

$$V^*_{\text{off}}(n,k) - V^*_{\text{on}}(n,k) \leq V^*_{\text{off}}(n,k) - V^{\text{br}}_{\text{on}}(n,k) \leq a_1 M \qquad \text{for all } (n,k) \in \mathcal{T}.$$

The BR policy, although achieving bounded regret, is *not* the optimal policy. Considering the optimality equation (C.2) developed in Appendix C, one can see that with two periods to go ($\ell = 2$) and one unit of budget ($k = 1$), the optimal action is to take any (and only) values $a_j \geq h_2(1) = g_1(1) - g_1(0) = \mathbb{E}[X_1]$. In this state, the BR policy will instead take all values above the median. Of course, although the BR policy makes some

Figure 4. Ratio Sample Paths



Notes. The figure shows sample paths of the ratio $R^\pi_t = \frac{1}{n-t}(k - \sum_{s\in[t]}\sigma^\pi_s)$ when π is the optimal dynamic-programming (DP) policy or the BR policy. The ability distribution is uniform on $\mathcal{A} = \{0.20, 0.65, 1.10, 1.55, 2.00\}$. The two plots have common random numbers but different initial conditions. On the left, we take $k = 300$, and on the right, we have $k = 340$. In both instances, the initial ratio k/n is within the orbit of $T_2 = 0.30$. When $k/n = 0.30 = T_2$ (left) the BR R^{br}_t stays within the orbit of T_2 for most time periods. When $k/n = 0.34 > T_2$ (right), the BR is first pulled toward the threshold T_2, and then it stays within its orbit until its escape time toward the end of the horizon. Despite just being on a sample path, the ratios R^*_t and R^{br}_t are remarkably close to each other.

mistakes when approaching the horizon's end, Corollary 1 is evidence that it mostly does the right thing. Figures 5 and 6 are numerical evidence for the performance of the BR policy.

Proof of Corollary 1. It suffices to prove that the BR policy satisfies the sufficient conditions stated in Proposition 2. By Theorem 2, the BR policy has the associated stopping time τ in (20) that satisfies condition (iv) in the proposition. Conditions (i)–(iii) are verified by the following sample-path argument.

If $n < 2\delta^{-1} + 1$, then the three conditions are satisfied immediately by choosing M to be a suitable constant. Otherwise, if $n \geq 2\delta^{-1} + 1$, we note that $T_j = \bar{F}(a_j) + \frac{1}{2} f_j = \bar{F}(a_{j+1}) - \frac{1}{2} f_j$ for all $j \in \{2, \ldots, m\}$, so the definition (5) tells us that if $k/n \in [T_j, T_{j+1})$, then $j_0(n,k) = j$ for all $j \in [m]$. Thus, if j is the index such that $k/n \in [T_j, T_{j+1})$, then j is the "action" index identified in Proposition 2, and we verify the three conditions by distinguishing the case $\{\tau_0 < n - 2\delta^{-1} - 1\}$ from $\{\tau_0 = n - 2\delta^{-1} - 1\}$.

As argued earlier, in the event $\{\tau_0 < n - 2\delta^{-1} - 1\}$, the index $j(\tau_0) \in \{j, j+1\}$. In turn, for all $t < \tau_0$, the BR policy selects $X_{t+1} \geq a_j$ and skips all $X_{t+1} \leq a_{j+1}$. If $j(\tau_0) = j$, then for time indices $t \in [\tau_0, \tau)$, all values a_{j-1} and greater are selected, and all values a_{j+1} and smaller are skipped. If $j(\tau_0) = j+1$, then on $t \in [\tau_0, \tau)$, all values a_j and greater are selected, and all those smaller than or equal to a_{j+2} are skipped. Thus, in the event $\{\tau_0 < n - 2\delta^{-1} - 1\}$, we have

$$\sum_{i \in [j-1]} S_i^{\mathrm{br},\tau} = \sum_{i \in [j-1]} Z_i^{\tau}, \tag{21}$$

where, recall, $S_i^{\mathrm{br},\tau}$ is the number of a_i candidates selected by time τ. Furthermore, we have the following:

1. If $j(\tau_0) = j$, then

$$S_j^{\mathrm{br},\tau} = \min\left\{ Z_j^{\tau}, k - K_{\tau} - \sum_{i \in [j-1]} Z_i^{\tau} \right\} = \min\left\{ Z_j^{\tau}, \left(k - \sum_{i \in [j-1]} Z_i^{\tau} \right)_+ - K_{\tau} \right\} \quad \text{and} \quad S_{j+1}^{\mathrm{br},\tau} = 0. \tag{22}$$

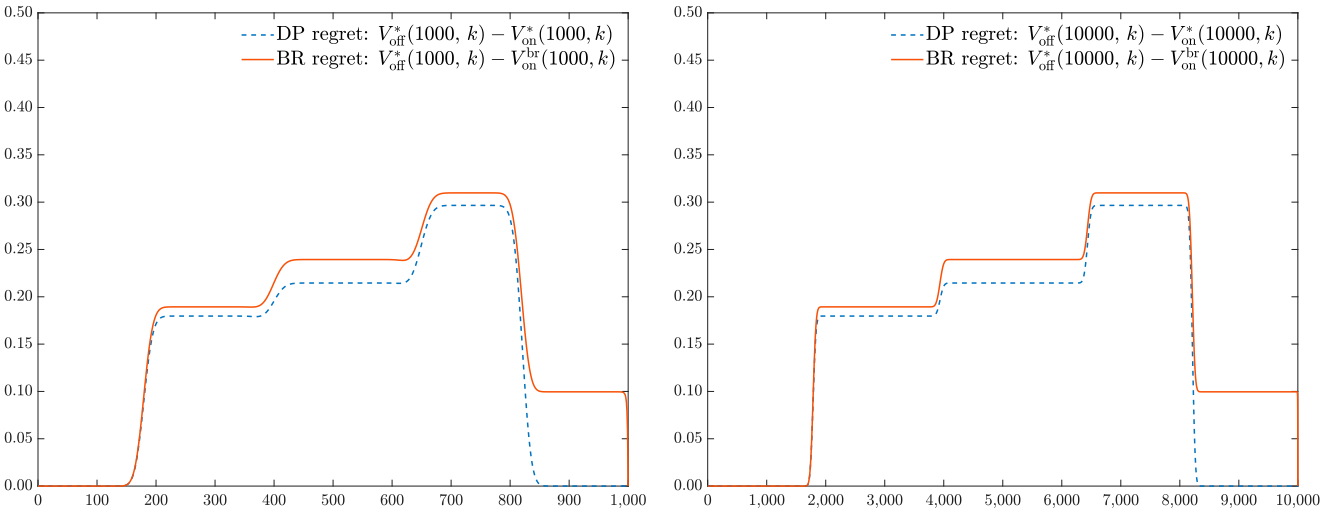
Here the left equality holds because out of the total budget used by time τ, which is given by $k - K_{\tau} = \sum_{t \in [\tau]} \sigma_t^{\mathrm{br}}$, the quantity $\sum_{i \in [j-1]} Z_i^{\tau}$ is allocated to values larger than or equal to a_{j-1} and the remaining to a_j.

2. If $j(\tau_0) = j+1$, then we have that

$$S_j^{\mathrm{br},\tau} = Z_j^{\tau} \quad \text{and} \quad S_{j+1}^{\mathrm{br},\tau} = \min\left\{ Z_{j+1}^{\tau}, k - K_{\tau} - \sum_{i \in [j]} Z_i^{\tau} \right\} = \min\left\{ Z_{j+1}^{\tau}, \left(k - \sum_{i \in [j]} Z_i^{\tau} \right)_+ - K_{\tau} \right\}, \tag{23}$$

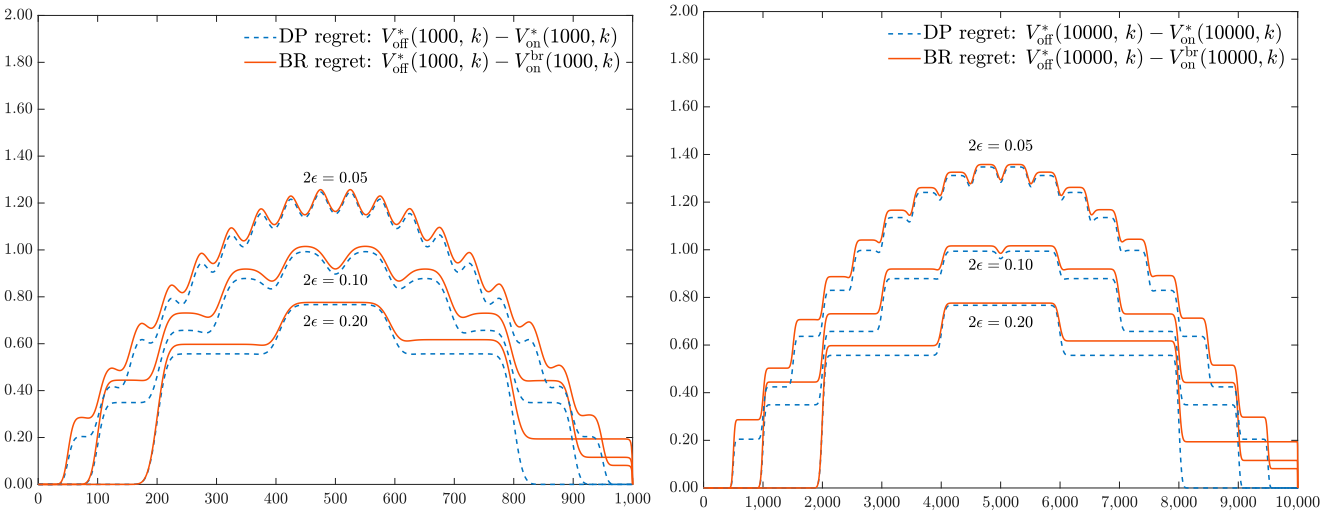
where the right equality holds for reasons that are similar to those given just below (22).

Figure 5. Dynamic-Programming and BR Regrets



Notes. The figure displays the dynamic-programming (DP) and BR regrets for the five-point distribution on $\mathcal{A} = \{0.2, 0.5, 0.7, 0.8, 1.0\}$ with probability mass function $(f_5, f_4, f_3, f_2, f_1) = (\frac{5}{28}, \frac{5}{28}, \frac{7}{28}, \frac{6}{28}, \frac{5}{28})$. We fix n to either $n = 1{,}000$ (left) or $n = 10{,}000$ (right) and vary the initial budgets k. The regret is small and piecewise constant. The "jumps" in the regret function match the jump points of the distribution.

Figure 6. Dynamic-Programming and BR Regrets for Uniform Distributions



Notes. The figure displays the dynamic-programming (DP) and BR regrets for uniform distributions over $[0.2, 0.2+\Delta, 0.2+2\Delta, \ldots, 2.0-\Delta, 2.0]$, where $\Delta = (2.0-0.2)2\epsilon/(1-2\epsilon)$ for $2\epsilon \in \{0.05, 0.10, 0.20\}$. As ϵ shrinks, the cardinality of the support grows. We fix $n = 1,000$ (left) and $n = 10,000$ (right) and vary the recruiting budget k. As n grows, the regret approaches a piecewise constant function. The optimality gap of the BR policy is, regardless, very small.

By combining the observations in (22) and (23), we find, in the event $\{\tau_0 < n - 2\delta^{-1} - 1\}$, that

$$S_j^{\text{br},\tau} \geq \min\left\{\left(k - \sum_{i\in[j-1]} Z_i^\tau\right)_+, Z_j^\tau\right\} - K_\tau = \mathfrak{S}_j^\tau - K_\tau \tag{24}$$

and

$$S_{j+1}^{\text{br},\tau} \geq \min\left\{\left(k - \sum_{i\in[j]} Z_i^\tau\right)_+, Z_{j+1}^\tau\right\} - K_\tau = \mathfrak{S}_{j+1}^\tau - K_\tau, \tag{25}$$

where we use the fact that if x, y, and z are nonnegative numbers, then $\min(x-y, z) \geq \min(x,z) - y$.

In the event $\{\tau_0 = n - 2\delta^{-1} - 1\}$, we have that $\tau = \tau_0$, and all values greater than or equal to a_j are selected and all values lower than or equal to a_{j+1} are skipped so that

$$\sum_{i\in[j-1]} S_i^{\text{br},\tau} = \sum_{i\in[j-1]} Z_i^\tau \qquad \text{and} \qquad S_j^{\text{br},\tau} = \mathfrak{S}_j^\tau, \tag{26}$$

and

$$0 = S_{j+1}^{\text{br},\tau} = \left(k - \sum_{i\in[j]} Z_i^\tau\right)_+ - K_\tau \geq \min\left\{Z_{j+1}^\tau, \left(k - \sum_{i\in[j]} Z_i^\tau\right)_+\right\} - K_\tau = \mathfrak{S}_{j+1}^\tau - K_\tau. \tag{27}$$

Finally, because the BR policy selects all remaining values as soon as there is a $t \in [n]$ such that $K_t \geq n - t$, we have that $K_\tau \leq n - \tau$ and, consequently, that

$$-\mathbb{E}[K_\tau] \geq \mathbb{E}[\tau] - n \geq -M, \tag{28}$$

where the last inequality follows from Theorem 2.

If we now recall the estimates (21), (24), and (25), which hold in the event that $\{\tau_0 < n - 2\delta^{-1} - 1\}$, and the relations (26) and (27), which are satisfied on $\{\tau_0 = n - 2\delta^{-1} - 1\}$, take expectations, and recall the bound (28), we see that the sufficient conditions (i)–(iii) in Proposition 2 are all satisfied. □

4.1. An Alternative to the BR Policy: The Adaptive-Index Policy

In closely related work, Wu et al. (2015) offer an elegant adaptive-index policy that we revisit here. Given the deterministic relaxation (15), we re-solve in each time period the deterministic problem

$$DR(n-t, K_t) = \varphi(\mathbb{E}[Z_1^{n-t}], \ldots, \mathbb{E}[Z_m^{n-t}], K_t),$$

and recall from (16) that the optimal solution is given by $s^*_{j,t} = \min\{\mathbb{E}[Z_j^{n-t}], (K_t - \sum_{i\in[j-1]}\mathbb{E}[Z_i^{n-t}])_+\} = \min\{f_j(n-t), (K_t - \bar{F}(a_j)(n-t))_+\}$ for all $j \in [m]$.

Then we construct the adaptive (reoptimized) index policy by mimicking the solution of this optimization problem. Specifically, if $s^*_{j,t} = f_j(n-t)$ and the candidate inspected at time $t+1$ has ability a_j, then that candidate is selected. Otherwise, if $s^*_{j,t} = K_t - \bar{F}(a_j)(n-t) > 0$, then an arriving a_j candidate is selected with probability

$$p_{j',t+1} = \frac{K_t/(n-t) - \bar{F}(a_{j'})}{f_{j'}}.$$

Finally, if $s^*_{j,t} = 0$, then an arriving a_j candidate is rejected.

This policy induces a nice martingale structure. Using the notation introduced in Section 2, we let $\sigma_1^{\mathrm{ai}}, \sigma_2^{\mathrm{ai}}, \ldots, \sigma_n^{\mathrm{ai}}$ be the sequence of decisions of the adaptive-index policy and let $K_t^{\mathrm{ai}} = k - \sum_{s\in[t]}\sigma_s^{\mathrm{ai}}$ be the associated remaining budget process for all $t \in [n]$ and with $K_0^{\mathrm{ai}} = k$. In the statement and in the proof of the next proposition, we use the standard notation $a \wedge b = \min\{a, b\}$.

Proposition 3. *Let* $\tau_1 = \inf\{t \in [n] : K_t^{\mathrm{ai}}/(n-t) > 1\}$. *Then the stopped adaptive-index ratio process* $\{R^{\mathrm{ai}}_{t\wedge\tau_1} = \frac{K^{\mathrm{ai}}_{t\wedge\tau_1}}{n-(t\wedge\tau_1)} : t \in [n]\}$ *is a martingale.*

Proof. Because $0 \le R^{\mathrm{ai}}_{t\wedge\tau_1} \le 1$, it is trivially true that $\mathbb{E}[|R^{\mathrm{ai}}_{t\wedge\tau_1}|] < \infty$ for all $t \in [n]$. Next, if $\mathcal{F}_t$ is the σ-field generated by the random variables $\sigma_1^{\mathrm{ai}}, \ldots, \sigma_t^{\mathrm{ai}}$, then we have that

$$\mathbb{E}[R^{\mathrm{ai}}_{(t+1)\wedge\tau_1} - R^{\mathrm{ai}}_{t\wedge\tau_1}|\mathcal{F}_t] = \mathbb{E}[(R^{\mathrm{ai}}_{(t+1)\wedge\tau_1} - R^{\mathrm{ai}}_{t\wedge\tau_1})\mathbb{1}(\tau_1 > t)|\mathcal{F}_t] = \mathbb{1}(\tau_1 > t)\mathbb{E}\left[\frac{K_t^{\mathrm{ai}} - \sigma_{t+1}^{\mathrm{ai}}}{n-(t+1)} - \frac{K_t^{\mathrm{ai}}}{n-t}\,\middle|\,\mathcal{F}_t\right],$$

where $\sigma_{t+1}^{\mathrm{ai}} = 1$ with probability $\frac{K_t^{\mathrm{ai}}}{n-t} \wedge 1$ and 0 otherwise. Hence,

$$\mathbb{E}[R^{\mathrm{ai}}_{(t+1)\wedge\tau_1} - R^{\mathrm{ai}}_{t\wedge\tau_1}|\mathcal{F}_t] = \left(\frac{K_t^{\mathrm{ai}} - \frac{K_t^{\mathrm{ai}}}{n-t} \wedge 1}{n-(t+1)} - \frac{K_t^{\mathrm{ai}}}{n-t}\right)\mathbb{1}(\tau_1 > t) = 0,$$

where we use the fact that $\mathbb{1}(\tau_1 > t)$ implies $K_t^{\mathrm{ai}}/(n-t) \le 1$. □

Because of this martingale structure, the adaptive-index ratio $K_t^{\mathrm{ai}}/(n-t)$ remains "close" to k/n so that this policy, like the BR policy, is careful in utilizing its budget and does not run out of it until (almost) the horizon's end. Furthermore, this martingale property guarantees bounded regret when the initial ratio k/n is safely far from the masses of the discrete distribution (see also Appendix D and Wu et al. 2015).

In general, however, spending the budget at the right "rate" is not sufficient. It is also important that the budget is spent on the right candidates, and the symmetric martingale structure is too weak for that purpose. When initialized, for example, at $k/n = \bar{F}(a_j)$ for some j, the martingale spends an equal amount of time below and above $\bar{F}(a_j)$, and the reoptimized index policy takes the values a_j and a_{j+1} in equal proportions. A good policy should start selecting a_{j+1} values only after it has selected all (or most) a_j values first. Figure 7 gives an illustration of this phenomenon.

4.2. Proof of Theorem 2

Given the deviation process

$$Y_u = K_{\tau_0+u} - T_{j(\tau_0)}(n - \tau_0 - u), \qquad u \in \{0, 1, \ldots n - \tau_0\},$$

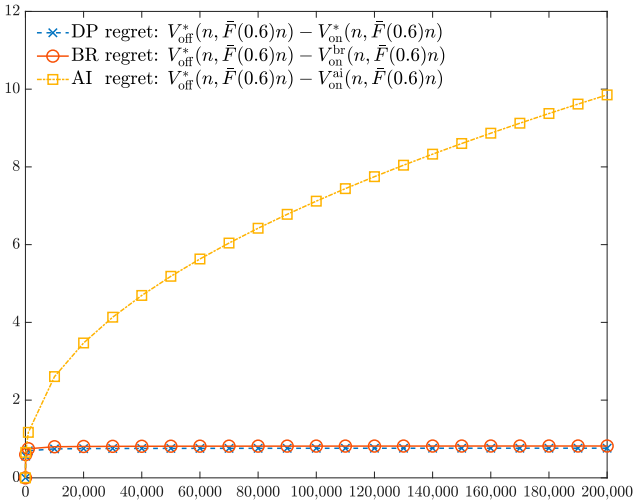
the key step in the proof of Theorem 2 is the derivation of an exponential tail bound for the random variable $|Y_u|$ for each $u \in \{0, 1, \ldots, n - \tau_0\}$. As before, we take $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$.

Proposition 4 (Exponential Tail Bound). *Fix* $0 < \delta < \epsilon$, $c = e^2 - 3$, *and* $0 < \eta < (\epsilon - \delta)/c$. *Then there is a constant* $M \equiv M(\epsilon)$ *such that for all* $0 \le u \le n - \tau_0$, *we have the exponential tail bound*

$$\mathbb{P}(|Y_u| \ge \delta(n - \tau_0 - u)\,|\,\widehat{\mathcal{F}}_0) \le 2\exp\left\{-\frac{\eta\delta}{2}(n - \tau_0)\right\} + M\exp\{-\eta\delta(n - \tau_0 - u)\}.$$

Proof. If $\tau_0 = n - 2\delta^{-1} - 1$, then the statement is trivial. Otherwise, for $\tau_0 < n - 2\delta^{-1} - 1$, the proof is an application—using the mean-reversal property of the BR policy—of the tail bound of Hajek (1982) to the two processes $\{Y_u : 0 \le u \le n - \tau_0\}$ and $\{-Y_u : 0 \le u \le n - \tau_0\}$.

Figure 7. Dynamic-Programming (DP), Budget-Ratio (BR), and Adaptive-Index (AI) Regrets



Notes. The graph displays the regrets of different policies when the ability distribution is uniform on $\mathscr{A} = \{0.2, 0.4, 0.6, \ldots, 1, 1.2, 1.4, \ldots, 2.0\}$ as a function of the horizon length n. In the chart, we vary n from 0 to 200,000 and keep the initial budget at $\frac{k}{n} = \bar{F}(0.6)$ (a mass point of the distribution). Whereas the optimal and BR policies achieve bounded regret, the regret of the adaptive-index policy grows with the problem size n.

We begin with the former. For any $a \geq 0$, we have that $\mathbb{1}(Y_u > a) = \mathbb{1}(Y_u > a, K_{\tau_0+u} > 0)$ for all $0 \leq u \leq n - \tau_0$, so that, by replacing $f_{j(\tau_0)}$ with $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$ in (17), we obtain the condition

$$\mathbb{E}[\,Y_{u+1} - Y_u | \widehat{\mathscr{F}}_u]\mathbb{1}(Y_u > a) \leq -\epsilon \mathbb{1}(Y_u > a) \qquad \text{for all } a \geq 0 \text{ and } 0 \leq u \leq n - \tau_0. \tag{HC1}$$

Because, by definition,

$$|Y_{u+1} - Y_u| = |-\sigma^{\text{br}}_{\tau_0+u+1} - T_{j(\tau_0)}| \leq 2 \qquad \text{for all } 0 \leq u < n - \tau_0,$$

we also have the condition

$$\mathbb{E}[\exp\{\lambda |Y_{u+1} - Y_u|\} | \widehat{\mathscr{F}}_u] \leq e^{2\lambda} \qquad \text{for all } \lambda > 0. \tag{HC2}$$

Conditions (HC1) and (HC2) (with $\lambda = 1$) imply, by Hajek (1982, lemma 2.1) that for η, ρ satisfying

$$0 < \eta < (\epsilon - \delta)/c \qquad \text{and} \qquad \rho = 1 - \eta(\epsilon - \eta c), \tag{29}$$

we have

$$\mathbb{E}[\exp\{\eta(Y_{u+1} - Y_u)\} | \widehat{\mathscr{F}}_u]\mathbb{1}(Y_u > a) \leq \rho \qquad \text{and} \qquad \mathbb{E}[\exp\{\eta(Y_{u+1} - a)\} | \widehat{\mathscr{F}}_u]\mathbb{1}(Y_u \leq a) \leq e^2,$$

for all $a \geq 0$ and all $0 \leq u < n - \tau_0$. Hajek (1982, theorem 2.3, equation 2.8) then gives

$$\mathbb{P}(Y_u \geq \delta(n - \tau_0 - u) | \widehat{\mathscr{F}}_0) \leq \rho^u \exp\{\eta Y_0 - \eta\delta(n - \tau_0 - u)\} + \frac{e^2}{1 - \rho}\exp\{\eta a - \eta\delta(n - \tau_0 - u)\},$$

for all $a \geq 0$ and $0 \leq u \leq n - \tau_0$. Taking $a = 0$ and using the fact that and $Y_0 \leq |Y_0| = |K_{\tau_0} - T_{j(\tau_0)}(n - \tau_0)| \leq \frac{\delta}{2}(n - \tau_0)$ as well as the standard inequality $\rho = 1 - \eta(\epsilon - \eta c) \leq \exp\{-\eta(\epsilon - \eta c)\}$, we finally have the upper bound

$$\mathbb{P}(Y_u \geq \delta(n - \tau_0 - u) | \widehat{\mathscr{F}}_0) \leq \exp\left\{-\frac{\eta\delta}{2}(n - \tau_0) - \eta u(\epsilon - \eta c - \delta)\right\} + \frac{e^2}{1 - \rho}\exp\{-\eta\delta(n - \tau_0 - u)\}.$$

The choice of η in (29) tells us that $\epsilon - \eta c - \delta > 0$, so we can drop the second term in the first exponent on the right-hand side. By setting $M \equiv M(\epsilon) = \frac{e^2}{1-\rho}$, we then have

$$\mathbb{P}(Y_u \geq \delta(n - \tau_0 - u) | \widehat{\mathscr{F}}_0) \leq \exp\left\{-\frac{\eta\delta}{2}(n - \tau_0)\right\} + M\exp\{-\eta\delta(n - \tau_0 - u)\}. \tag{30}$$

The analysis of the sequence $\{-Y_u : 0 \le u \le n-\tau_0\}$ follows a similar logic. For any $a \ge 0$,

$$\begin{aligned}\mathbb{E}[-Y_{u+1}+Y_u|\widehat{\mathcal{F}}_u]\mathbb{1}(-Y_u>a) = &-\mathbb{E}[Y_{u+1}-Y_u|\widehat{\mathcal{F}}_u]\mathbb{1}(Y_u<-a, K_{\tau_0+u}>0)\\ &-\mathbb{E}[Y_{u+1}-Y_u|\widehat{\mathcal{F}}_u]\mathbb{1}(Y_u<-a, K_{\tau_0+u}=0),\end{aligned}$$

so by (18) and (19), we have

$$\mathbb{E}[-Y_{u+1}+Y_u|\widehat{\mathcal{F}}_u]\mathbb{1}(-Y_u>a) \le -\tfrac{1}{2}f_{j(\tau_0)}\mathbb{1}(Y_u<-a, K_{\tau_0+u}>0) - T_{j(\tau_0)}\mathbb{1}(Y_u<-a, K_{\tau_0+u}=0),$$

for all $a \ge 0$. In the event $\{Y_u < -a, K_{\tau_0+u}=0\}$, we must have that $j(\tau_0)>1$. Otherwise, if $j(\tau_0)=1$, $T_{j(\tau_0)}=0$, and $Y_u = K_{\tau+u} = 0 \ge -a$ by definition. In particular, $T_{j(\tau_0)} \ge \epsilon$ in this event, and it follows that

$$\mathbb{E}[-Y_{u+1}+Y_u|\widehat{\mathcal{F}}_u]\mathbb{1}(-Y_u>a) \le -\epsilon\mathbb{1}(-Y_u>a), \tag{$\widetilde{\text{HC1}}$}$$

for all $a \ge 0$. It is then easily verified that

$$\mathbb{E}\big[\exp\{\lambda|Y_{u+1}-Y_u|\}\,|\widehat{\mathcal{F}}_u\big] \le e^{2\lambda} \qquad \text{for all } \lambda>0. \tag{$\widetilde{\text{HC2}}$}$$

As before, Hajek (1982, lemma 2.1 and theorem 2.3) gives—with $c=e^2-3$, $0<\eta\le(\epsilon-\delta)/c$, $\rho=1-\eta(\epsilon-\eta c)$, and $M \equiv M(\epsilon) = \frac{e^2}{1-\rho}$ —that

$$\mathbb{P}(-Y_u \ge \delta(n-\tau_0-u)|\widehat{\mathcal{F}}_0) \le \exp\left\{-\frac{\eta\delta}{2}(n-\tau_0)\right\} + M\exp\{-\eta\delta(n-\tau_0-u)\}. \tag{31}$$

The statement of the lemma is now the combination of (30) and (31) with $M \equiv M(\epsilon) = \frac{2e^2}{1-\rho}$. □

The exponential tail bound in Proposition 4 goes a long way for the proof of Theorem 2, which follows next.

Proof of Theorem 2. By definition, for $\tau_0 < t \le n - 2\delta^{-1} - 1$, we have that

$$\tau \le t \text{ if and only if } \left|\frac{K_{\tau_0+u}}{n-\tau_0-u} - T_{j(\tau_0)}\right| > \delta, \text{ for some } 0 \le u \le t-\tau_0,$$

or, represented in terms of Y_u,

$$\tau \le t \qquad \text{if and only if} \qquad \sum_{u\in[t-\tau_0]} \mathbb{1}(|Y_u| > \delta(n-\tau_0-u)) \ge 1.$$

We will represent $\mathbb{E}[\tau]$ as a sum of the tail probabilities

$$\mathbb{P}(\tau \le t|\widehat{\mathcal{F}}_0) = \mathbb{P}\left(\sum_{u\in[t-\tau_0]} \mathbb{1}(|Y_u| > \delta(n-\tau_0-u)) \ge 1|\widehat{\mathcal{F}}_0\right), \qquad \text{for } \tau_0 < t \le n-2\delta^{-1}-1,$$

which, by Markov's inequality, satisfy the bounds

$$\mathbb{P}(\tau \le t|\widehat{\mathcal{F}}_0) \le \sum_{u\in[t-\tau_0]} \mathbb{P}(|Y_u| > \delta(n-\tau_0-u)|\widehat{\mathcal{F}}_0), \qquad \text{for } \tau_0 < t \le n-2\delta^{-1}-1.$$

By integrating the exponential tail bound in Proposition 4 for $0 \le u \le t-\tau_0$ (recall that $\tau_0 < t \le n-2\delta^{-1}-1 < n$), we obtain

$$\mathbb{P}(\tau \le t|\widehat{\mathcal{F}}_0) \le 2(n-\tau_0)\exp\left\{-\frac{\eta\delta}{2}(n-\tau_0)\right\} + M\exp\{-\eta\delta(n-t)\}, \tag{32}$$

for some constant $M \equiv M(\epsilon)$. Because $\tau > \tau_0$ by definition, then $\mathbb{P}(\tau > t|\widehat{\mathcal{F}}_0) = 1$ for all $t \le \tau_0$, and we also have that

$$\mathbb{E}[\tau|\widehat{\mathcal{F}}_0] \ge \sum_{t=0}^{\tau_0} \mathbb{P}(\tau > t|\widehat{\mathcal{F}}_0) + \sum_{t=\tau_0+1}^{n-2/\delta-1} \mathbb{P}(\tau > t|\widehat{\mathcal{F}}_0) \ge n - 2\delta^{-1} - 1 - \sum_{t=\tau_0+1}^{n-2/\delta-1} \mathbb{P}(\tau \le t|\widehat{\mathcal{F}}_0).$$

Integrating (32) then gives us another constant $M \equiv M(\epsilon)$ such that

$$\mathbb{E}[\tau|\widehat{\mathscr{F}}_0] \geq n - 2(n-\tau_0)^2 \exp\{-\tfrac{\eta^\delta}{2}(n-\tau_0)\} - M.$$

The middle summand on the right-hand side is uniformly bounded, so in summary we have that

$$\mathbb{E}[\tau|\widehat{\mathscr{F}}_0] \geq n - M,$$

for some constant $M \equiv M(\epsilon) < \infty$, and the proof of the theorem follows after one takes total expectations. □

5. The Square-Root Regret of Nonadaptive Policies

A nonadaptive policy π is an online feasible policy that is characterized by a probability matrix $\{p_{j,t} : j \in [m] \text{ and } t \in [n]\}$. The entry $p_{j,t}$ represents the probability of selecting the candidate inspected at time t given that the candidate's ability is $X_t = a_j$ and that the recruiting budget remaining at time t is nonzero. Formally, given $\pi = \{p_{j,t} : j \in [m] \text{ and } t \in [n]\}$, let

$$B_t = \sum_{j\in[m]} \mathbb{1}(U_t \leq p_{j,t}, X_t = a_j) \qquad \text{for } t \in [n],$$

and note that the sequence $B_1, B_2, \ldots, B_n$ is a sequence of $\mathscr{F}_t$-measurable, independent Bernoulli random variables with success probabilities $q_1, q_2, \ldots, q_n$ such that

$$\mathbb{E}[B_t \,|\, X_t = a_j] = p_{j,t} \qquad \text{and} \qquad q_t = \mathbb{E}[B_t] = \sum_{j\in[m]} p_{j,t} f_j.$$

The nonadaptive policy π selects candidates until it runs out of budget or reaches the end of the horizon, that is, up to the stopping time $\nu \equiv \nu(\pi)$ given by

$$\nu = \min\left\{r \geq 1 : \sum_{t\in[r]} B_t \geq k \text{ or } r \geq n\right\}.$$

The expected total ability accrued by the nonadaptive policy π is then given by

$$V^\pi_{\text{on}}(n,k) = \sum_{j\in[m]} a_j \mathbb{E}\left[\sum_{t\in[\nu]} B_t \mathbb{1}(X_t = a_j)\right],$$

and $V^*_{\text{na}}(n,k) = \sup_{\pi\in\Pi_{\text{na}}} V^\pi_{\text{on}}(n,k)$ is the performance of the best nonadaptive policy.

An intuitive nonadaptive policy is the index policy $\text{id} = \{\mathfrak{p}_{j,t} : j \in [m] \text{ and } t \in [n]\}$, which takes its probabilities from the solution (16) to the deterministic relaxation (15). If the residual budget at time t is nonzero, and if $j_{\text{id}} \in [m]$ is the index such that $\bar{F}(a_{j_{\text{id}}}) \leq k/n < \bar{F}(a_{j_{\text{id}}+1})$, then the index policy selects an arriving a_j candidate with probability

$$\mathfrak{p}_{j,t} = \frac{s^*_j}{\mathbb{E}[Z^n_j]} = \frac{s^*_j}{nf_j} = \begin{cases} 1 & \text{if } j \leq j_{\text{id}} - 1, \\ \dfrac{k/n - \bar{F}(a_{j_{\text{id}}})}{f_{j_{\text{id}}}} & \text{if } j = j_{\text{id}}, \\ 0 & \text{if } j \geq j_{\text{id}} + 1. \end{cases} \tag{33}$$

In the main result of this section, we prove that, for a large range of (n,k) pairs, the regret of nonadaptive policies is generally of the order of $\sqrt{n}$. Viewed in the context of Proposition 2, nonadaptive policies violate property (iv): they run out of budget too early. The proof relies on this "greediness."

To build intuition, consider the problem instance with ability distribution that is uniform on $\mathscr{A} = \{a_3, a_2, a_1\}$ and with $n = 2k$. For each $t \in [n]$, the index policy *would* select a_1 candidates with probability $\mathfrak{p}_{1,t} = 1$, a_2 candidates with probability $\mathfrak{p}_{2,t} = 1/2$, and no a_3 candidates. The number of selections that the index policy would make by time t is then a binomial random variable with t trials and success probability equal to $k/n = 1/2$. For $t = \lfloor n - \sqrt{n} \rfloor$, this random variable has mean approximately equal to $\frac{1}{2}(n - \sqrt{n}) = (1 - \frac{1}{\sqrt{n}})\frac{n}{2}$ and standard deviation approximately equal to $\frac{1}{4}\sqrt{n - \sqrt{n}}$. In this case, the recruiting budget $k = n/2$ is, in the limit, 2 standard deviations from the mean so that, with nonnegligible probability, the policy would have exhausted all its budget by time $n - \sqrt{n}$ and missed $\Theta(\sqrt{n})$ opportunities to select a_1 values. The index policy, however,

did select many a_2 values up to this time t. The BR policy—and the dynamic-programming policy—would exchange those a_2 with a_1 and attain a regret that does not grow with n.

Theorem 3 (The Regret of Nonadaptive Policies). *For $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$, suppose that $(f_1+\epsilon)n \le k \le (1-f_m-\epsilon)n$. Then there is a constant $M \equiv M(\epsilon, m, a_1, \ldots, a_m)$ such that*

$$M\sqrt{n} \le V^*_{\text{off}}(n,k) - V^*_{\text{na}}(n,k).$$

As one might expect, the nonadaptive (time-homogeneous) index policy in (33) already achieves this order of magnitude and, in this sense, is representative of the performance of general nonadaptive policies.

Lemma 3 (The Regret of the Nonadaptive Index Policy). *The nonadaptive index policy $\text{id} = \{\mathfrak{p}_{j,t} : j \in [m] \text{ and } t \in [n]\}$ has a $\sqrt{n}$ regret; that is, for any $\varepsilon \in (0,1)$ and all pairs (n,k) such that $\varepsilon \le k/n$, we have*

$$V^*_{\text{off}}(n,k) - V^{\text{id}}_{\text{na}}(n,k) \le DR(n,k) - V^{\text{id}}_{\text{na}}(n,k) \le \varepsilon^{-1} a_1 \sqrt{n}.$$

The conditions of Theorem 3 are not necessary, but certain budget ranges do have to be excluded. In the small-budget range with $k \le (f_1 - \epsilon)n$, for example, the regret is in fact constant. The offline solution mostly takes a_1 values. The nonadaptive policy $\widehat{\pi}$ that has $p_{1,t} \equiv 1$ for all $t \in [n]$ and $p_{j,t} \equiv 0$ for all $j \ne 1$ and all $t \in [n]$ will achieve a constant regret. Interestingly, in this same range, the index policy may not be as aggressive. For instance, if $k = (f_1 - \epsilon)n$, we see from (33) that the index policy sets $\mathfrak{p}_{1,t} = 1 - \epsilon/f_1$, spreading out the selection of a_1 values throughout the time horizon and having a regret of order $\sqrt{n}$.

Lemma 4 (Nonadaptive Regret: Small Budget). *For $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$, suppose that $k \le n(f_1 - \epsilon)$. Then one has that*

$$V^*_{\text{off}}(n,k) - V^*_{\text{na}}(n,k) \le \frac{a_2}{4\epsilon}.$$

Theorem 3 and Lemma 4 are nonexhaustive. There are ranges of k/n that are covered by neither. The purpose of these results is to show that although nonadaptivity is not universally bad (see Lemma 4), there is a nontrivial range of (k,n) pairs for which the regret grows like $\sqrt{n}$, whereas the BR and dynamic-programming policies achieve $O(1)$ regret.

In the proof of Theorem 3, we use three auxiliary lemmas, the first of which provides a lower bound for the overshoot of a centered Bernoulli random walk.

Lemma 5. *Let $B_1, B_2, \ldots, B_n$ be independent Bernoulli random variables with success probabilities $q_1, q_2, \ldots, q_n$, respectively. Let $N_n = \sum_{t\in[n]}\{B_t - q_t\}$, and let $\varsigma_n^2 = \text{Var}[N_n] = \sum_{t\in[n]} q_t(1-q_t)$. Then, for any $\Upsilon > 0$, there exist constants $\beta_1 \equiv \beta_1(\Upsilon)$ and $\beta_2 \equiv \beta_2(\Upsilon)$ such that*

$$\mathbb{E}[(N_n - \Upsilon\varsigma_n)_+] \ge \beta_1\varsigma_n - (2+3\sqrt{2}), \qquad \mathbb{E}[(-N_n - \Upsilon\varsigma_n)_+] \ge \beta_1\varsigma_n - (2+3\sqrt{2}) \tag{34}$$

and

$$\mathbb{E}[(N_n + \Upsilon\varsigma_n)_+^2] \le \beta_2\varsigma_n^2. \tag{35}$$

The second auxiliary lemma shows that any good nonadaptive policy must be a perturbation of the index policy. Specifically, if $s_j(\pi) = \sum_{t\in[n]} p_{j,t} f_j$ is the expected number of a_j candidates that policy π selects under infinite budget, then $s_j(\pi)$ is just a perturbation of s_j^*, the solution of the deterministic relaxation (15).

Lemma 6. *Fix $M < \infty$, $\epsilon = \frac{1}{2}\min\{f_m, \ldots, f_1\}$, and take any $(2\epsilon \min_{j\in[m-1]}|a_j - a_{j+1}|)^{-2}M^2 \le n$. If $\pi = \{p_{j,t} : j \in [m] \text{ and } t \in [n]\}$ is a nonadaptive policy such that*

$$DR(n,k) - V^{\pi}_{\text{on}}(n,k) \le M\sqrt{n},$$

then

$$s_j(\pi) = s_j^* \pm \left\{\min_{j\in[m-1]}|a_j - a_{j+1}|\right\}^{-1} M\sqrt{n} \qquad \text{for all } j \in [m].$$

For $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$ and $(f_1+\epsilon)n \le k \le (1-f_m-\epsilon)n$, Lemma 2 tells us that

$$nf_1 - \frac{1}{4\epsilon} \le \mathbb{E}[Z_1^n] - \mathbb{E}[(Z_1^n - k)_+] = \mathbb{E}[\mathfrak{S}_1^n] \qquad \text{and} \qquad \mathbb{E}[\mathfrak{S}_m^n] \le \frac{1}{4\epsilon}.$$

In words, when k is bounded away from both 0 and n, the offline algorithm selects—in expectation—all but a constant number of the highest values and at most a constant number of the lowest values. The optimal nonadaptive policy must do so as well. In particular, in most time periods the optimal nonadaptive policy should have $p_{1,t} = 1$ and $p_{m,t} = 0$ so that the marginal probability of selection q_t is safely bounded away from zero and one. The last auxiliary lemma makes this intuitive idea formal.

Lemma 7. *Let $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$, and suppose that $(f_1 + \epsilon)n \le k \le (1 - f_m - \epsilon)n$. Then there is a constant $M \equiv M(\epsilon, a_1, a_2, a_m) < \infty$ such that an optimal nonadaptive policy must satisfy*

$$\sum_{t\in[n]} \mathbb{1}\left\{q_t \ge \frac{f_1}{2}\right\} \ge n - M\sqrt{n} \quad \text{and} \quad \sum_{t\in[n]} \mathbb{1}\left\{1 - q_t \ge \frac{f_m}{2}\right\} \ge n - M\sqrt{n}. \tag{36}$$

Consequently,

$$\sum_{t\in[n]} \mathbb{1}\left\{q_t \ge \frac{f_1}{2}, 1 - q_t \ge \frac{f_m}{2}\right\} \ge n - 2M\sqrt{n},$$

and one has the lower bound

$$\varsigma^2(\pi) = \sum_{t\in[n]} q_t(1 - q_t) \ge \frac{f_1 f_m}{4}(n - 2M\sqrt{n}).$$

Proof of Theorem 3. For any nonadaptive policy $\pi = \{p_{j,t} : j \in [m] \text{ and } t \in [n]\}$ with associated stopping time $\nu = \min\{r \ge 1 : \sum_{t\in[r]} B_t \text{ or } r \ge n\}$, we have that policy π does not make any selection after time ν, and we also have that $Z_j^\nu \le Z_j^n$ for all $j \in [m]$. Thus, if we recall the linear program (2), use the monotonicity of $\varphi(z_1, \ldots, z_m, \cdot)$ in $(z_1, \ldots, z_m)$, and recall the equivalence (1), we obtain that

$$V_{\text{on}}^\pi(n,k) \le \mathbb{E}[\varphi(Z_1^\nu, \ldots, Z_m^\nu, k)] \le \mathbb{E}[\varphi(Z_1^n, \ldots, Z_m^n, k)] = V_{\text{off}}^*(n,k).$$

Furthermore, because ν is the time at which π exhausts the budget, we also have that

$$\begin{aligned} V_{\text{off}}^*(n,k) - V_{\text{on}}^\pi(n,k) &\ge \mathbb{E}[\varphi(Z_1^n, \ldots, Z_m^n, k)] - \mathbb{E}[\varphi(Z_1^\nu, \ldots, Z_m^\nu, k)] \\ &\ge a_1(\mathbb{E}[\min\{k, Z_1^n\}] - \mathbb{E}[\min\{k, Z_1^\nu\}]) \\ &\ge a_1\big(\mathbb{E}[(Z_1^n - Z_1^\nu)] - n\mathbb{1}(Z_1^n > k)]\big). \end{aligned}$$

For $\epsilon = \frac{1}{2}\min\{f_m, f_{m-1}, \ldots, f_1\}$ and $(f_1 + \epsilon)n \le k \le (1 - f_m - \epsilon)n$, Hoeffding's inequality (see, e.g., Boucheron et al. (2013), theorem 2.8) immediately tells us that there is a constant $M \equiv M(\epsilon) < \infty$ such that $n\mathbb{P}(Z_1^n > k) \le M$, and Wald's lemma gives us that

$$\mathbb{E}[Z_1^n - Z_1^\nu] = f_1\mathbb{E}[n - \nu], \tag{37}$$

so the proof of the theorem is complete if we can prove an appropriate lower bound for $\mathbb{E}[n - \nu]$. Lemma 3 tells us that for $f_1 + \epsilon \le k/n$, the index policy has regret that is bounded above $(f_1 + \epsilon)^{-1}a_1\sqrt{n}$, so it suffices to consider nonadaptive policies $\pi = \{p_{j,t}, j \in [m] \quad \text{and} \quad t \in [n]\}$ for which

$$DR(n,k) - V_{\text{on}}^\pi(n,k) \le (f_1 + \epsilon)^{-1}a_1\sqrt{n}.$$

If $(2\epsilon)^{-2}\bar{M}^2 \le n$ for $\bar{M} = \{\min_{j\in[m-1]}|a_j - a_{j+1}|(f_1 + \epsilon)\}^{-1}a_1$, Lemma 6 implies that

$$\sum_{t\in[n]} p_{j,t}f_j = s_j^* \pm \bar{M}\sqrt{n} \qquad \text{for all } j \in [m].$$

Because $\sum_{j\in[m]} s_j^* = k$, we also have that

$$\sum_{t\in[n]} q_t = \sum_{t\in[n]}\sum_{j\in[m]} p_{j,t}f_j = \sum_{j\in[m]} s_j^* \pm m\bar{M}\sqrt{n} = k \pm m\bar{M}\sqrt{n}. \tag{38}$$

Furthermore, because $0 \le B_t \le 1$ for all $t \in [n]$, we know that

$$\left(\sum_{t\in[n]} B_t - k\right)_+ = \sum_{t=\nu+1}^{n} B_t \le n-\nu, \qquad \text{so} \qquad \mathbb{E}\left[\left(\sum_{t\in[n]} B_t - k\right)_+\right] \le \mathbb{E}[n-\nu]. \tag{39}$$

With $\varsigma^2(\pi) = \sum_{t\in[n]} q_t(1-q_t)$, the estimate (38) implies that $k \le \sum_{t\in[n]} q_t + [m\bar{M}\,\varsigma(\pi)^{-1}\sqrt{n}\,]\varsigma(\pi)$, and Lemma 7 tells us that there is a constant $M \equiv M(\epsilon, m, a_1, \ldots, a_m)$ such that $m\bar{M}\,\varsigma(\pi)^{-1}\sqrt{n} \le M$. In turn, we also have that $k \le \sum_{t\in[n]} q_t + M\varsigma(\pi)$, so after we subtract $\sum_{t\in[n]} B_t$ on both sides, change signs, take the positive part, and recall (39), we obtain the lower bound

$$\left(\sum_{t\in[n]} (B_t - q_t) - M\varsigma(\pi)\right)_+ \le \left(\sum_{t\in[n]} B_t - k\right)_+ \le n-\nu.$$

The quantity on the left-hand side is a sum of centered independent Bernoulli random variables, so, when we take expectations, Lemma 5 tells us that there is a constant $\beta_1 \equiv \beta_1(\epsilon, m, a_1, \ldots, a_m)$ such that

$$\beta_1\varsigma(\pi) - (2+3\sqrt{2}\,) \le \mathbb{E}\left[\left(\sum_{t\in[n]} (B_t - q_t) - M\varsigma(\pi)\right)_+\right] \le \mathbb{E}[n-\nu].$$

Plugging this last estimate back into (37) gives us that $f_1(\beta_1\varsigma(\pi) - 2 - 3\sqrt{2}\,) \le \mathbb{E}[Z_1^n - Z_1^\nu]$, and the theorem then follows after one uses one more time the lower bound for $\varsigma(\pi)$ given in Lemma 7 and chooses $M \equiv M(\epsilon, m, a_1, \ldots, a_m)$ accordingly. □

6. Concluding Remarks

We have proved that in the multisecretary problem with independent candidate abilities drawn from a common finite-support distribution, the regret is constant and achievable by a multithreshold policy. In our model, the decision maker knows and makes crucial use of the distribution of candidate abilities. Two obvious extensions to consider are the problem instances in which the ability distribution is continuous and/or unknown to the decision maker.

Although one would like to think of the continuous distribution as a "limit" of discrete ones, our analysis does build to a great extent on this discreteness, and our bounds depend on the cardinality of the support. At this point, it is not clear whether bounded regret is achievable also with continuous distributions.

For the case of unknown distribution, we conjecture that, with a finite support, the regret should be logarithmic in n. Indeed, consider a "stupid" algorithm that uses the first $O(\log(n))$ steps to learn about the distribution (without concern for the objective) and, at the end of the learning period, computes the threshold and runs with the BR policy thereafter. Simple Chernoff bounds suggest that the likelihood of misestimation should be exponentially small. Coupling this with the fact that the performance of the BR policy (specifically that $\mathbb{E}[\tau] \ge n - M$) is insensitive to small perturbations to the thresholds leads to our conjecture.

Appendix A. Regret Lower Bound in Kleinberg's (2017) Example

Below we prove the regret lower bound in Lemma 1, in which the ability distribution has support $\mathcal{A} = (a_3, a_2, a_1)$ and probability mass function $(\frac{1}{2}+2\epsilon, 2\epsilon, \frac{1}{2}-4\epsilon)$. The argument shows that because of the special structure of this example, there are two events (of nonnegligible probability)—one being a perturbation of the other—where the offline sort makes very different choices (taking all or none of the a_2 values), but the optimal dynamic-programming policy can do well only in one or the other but not in both.

Proof of Lemma 1. In what follows, we will assume that ϵ is such that $\frac{1}{\epsilon^2}$ and $\frac{1}{2\epsilon^2}$ are integers so that we can set $n_\epsilon = \frac{1}{\epsilon^2}$ and $k_\epsilon = \frac{n_\epsilon}{2} = \frac{1}{2\epsilon^2}$. This makes the notation cleaner while not changing any of the arguments. Next, we introduce the following events:

$$A_\epsilon = \left\{\frac{1}{\epsilon} \le Z_2^{n_\epsilon} \le \min\left\{2Z_2^{n_\epsilon/2}, \frac{2}{\epsilon}\right\}\right\} = \left\{\frac{1}{2}\mathbb{E}[Z_2^{n_\epsilon}] \le Z_2^{n_\epsilon} \le \min\{2Z_2^{n_\epsilon/2}, \mathbb{E}[Z_2^{n_\epsilon}]\}\right\},$$

$$H_\epsilon = \left\{Z_1^{n_\epsilon} \ge k_\epsilon + \frac{2}{\epsilon}\right\} = \left\{Z_1^{n_\epsilon} \ge \mathbb{E}[Z_1^{n_\epsilon}] + \frac{6}{\epsilon}\right\}, \quad \text{and}$$

$$L_\epsilon = \left\{Z_1^{n_\epsilon} \le k_\epsilon - \frac{4}{\epsilon}\right\} = \left\{Z_1^{n_\epsilon} \le \mathbb{E}[Z_1^{n_\epsilon}]\right\}.$$

On the one hand, on the event H_ϵ—in particular, on $H_\epsilon \cap A_\epsilon$—the offline policy takes only the highest values; that is, it takes only a_1 values and exhausts the budget k_ϵ while doing so. On the other hand, on the event L_ϵ—in particular, on $L_\epsilon \cap A_\epsilon$—the offline policy also takes some lower values. In particular, it takes all the a_1 values, all the a_2 values, and some a_3 values to exhaust the budget. The events $H_\epsilon \cap A_\epsilon$ and $L_\epsilon \cap A_\epsilon$ are nontrivial because they happen with probability that is bounded away from zero: there are constants $\epsilon_0 > 0$ and $\alpha > 0$ such that $\mathbb{P}(H_\epsilon \cap A_\epsilon) \geq \alpha$ and $\mathbb{P}(L_\epsilon \cap A_\epsilon) \geq \alpha$ for all $\epsilon \leq \epsilon_0$. To see this, note that the standard deviation of $Z_1^{n_\epsilon}$ is on the order of $1/\epsilon$, so the central limit theorem implies that $\mathbb{P}(H_\epsilon), \mathbb{P}(L_\epsilon) \geq \alpha$ for all $\epsilon \leq \epsilon_0$. Another scaling argument shows that $\mathbb{P}(A_\epsilon) \geq \alpha > 0$ for all $\epsilon \leq \epsilon_0$ and for some redefined $\alpha > 0$. Because $Z_1^{n_\epsilon}$ and $Z_2^{n_\epsilon}$ are dependent, these observations and a standard condition argument that we omit imply that there is another $\alpha > 0$ such that $\mathbb{P}(H_\epsilon \cap A_\epsilon) \geq \alpha$ for all $\epsilon \leq \epsilon_0$. A similar analysis also gives that $\mathbb{P}(L_\epsilon \cap A_\epsilon)$ has the same property.

Next, we consider the random variable $\Lambda(n_\epsilon, k_\epsilon)$ that tracks the difference between the performance of the offline policy and that of the dynamic-programming policy

$$\Lambda(n_\epsilon, k_\epsilon) = a_1(\mathfrak{S}_1^{n_\epsilon} - S_1^{*,n_\epsilon}) + a_2(\mathfrak{S}_2^{n_\epsilon} - S_2^{*,n_\epsilon}) + a_3(\mathfrak{S}_3^{n_\epsilon} - S_3^{*,n_\epsilon}),$$

and we note that the expected value $\mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)] = V_{\text{off}}^*(n_\epsilon, k_\epsilon) - V_{\text{on}}^*(n_\epsilon, k_\epsilon)$ is the regret of the dynamic-programming policy. We now recall that $S_2^{*,n_\epsilon/2}$ is the number of a_2 selections by the optimal online policy up to time $n_\epsilon/2$, and we consider the event

$$C_\epsilon = \{S_2^{*,n_\epsilon/2} \geq Z_2^{n_\epsilon}/4\}.$$

On $C_\epsilon^c \cap L_\epsilon \cap A_\epsilon$, we have that

$$\mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)\mathbb{1}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon)] \geq \frac{a_2 - a_3}{4}\mathbb{E}[Z_2^{n_\epsilon}\mathbb{1}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon)] \geq \frac{a_2 - a_3}{4\epsilon}\mathbb{P}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon).$$

For each sequence $\{X_1, \ldots, X_{n_\epsilon}\} \in C_\epsilon \cap L_\epsilon \cap A_\epsilon$, we can now build a sequence in $C_\epsilon \cap H_\epsilon \cap A_\epsilon$ by keeping all the first $n_\epsilon/2$ values the same and replacing at most $6/\epsilon\, a_3$ values with a_1 values while keeping the number of a_2 values constant, and we call this resulting set D_ϵ. It is easy to see that because f_3 and f_1 are bounded away from zero, $\mathbb{P}(D_\epsilon) \geq \widehat{\alpha}\mathbb{P}(C_\epsilon \cap L_\epsilon \cap A_\epsilon))$ for some $\widehat{\alpha} > 0$ that does not depend on ϵ. Because the two sequences share the same first $n_\epsilon/2$ elements, we have that $S_2^{*,n_\epsilon/2} \geq Z_2^{n_\epsilon}/4$ also on D_ϵ. Then

$$\mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)\mathbb{1}(D_\epsilon)] \geq (a_1 - a_2)\mathbb{E}[S_2^{*,n_\epsilon/2}\mathbb{1}(D_\epsilon)] \geq \frac{a_1 - a_2}{4}\mathbb{E}[Z_2^{n_\epsilon}\mathbb{1}(D_\epsilon)] \geq \frac{a_1 - a_2}{4\epsilon}\widehat{\alpha}\mathbb{P}(C_\epsilon \cap L_\epsilon \cap A_\epsilon).$$

Because $\mathbb{P}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon) = \mathbb{P}(L_\epsilon \cap A_\epsilon) - \mathbb{P}(C_\epsilon \cap L_\epsilon \cap A_\epsilon)$, we have that $\max\{\mathbb{P}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon), \widehat{\alpha}\mathbb{P}(C_\epsilon \cap L_\epsilon \cap A_\epsilon)\} \geq \mathbb{P}(L_\epsilon \cap A_\epsilon)\min\{\widehat{\alpha}, 1\}$, so we finally obtain

$$\begin{aligned}
V_{\text{off}}^*(n_\epsilon, k_\epsilon) - V_{\text{on}}^*(n_\epsilon, k_\epsilon) &= \mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)] \\
&\geq \max\{\mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)\mathbb{1}(C^c \cap L_\epsilon \cap A_\epsilon)], \mathbb{E}[\Lambda(n_\epsilon, k_\epsilon)\mathbb{1}(D_\epsilon)]\} \\
&\geq \min\left\{\frac{a_2 - a_3}{4\epsilon}, \frac{a_1 - a_2}{4\epsilon}\right\}\max\{\mathbb{P}(C_\epsilon^c \cap L_\epsilon \cap A_\epsilon), \widehat{\alpha}\mathbb{P}(C_\epsilon \cap L_\epsilon \cap A_\epsilon)\} \\
&\geq \min\left\{\frac{a_2 - a_3}{4\epsilon}, \frac{a_1 - a_2}{4\epsilon}\right\}\min\{\widehat{\alpha}, 1\}\mathbb{P}(L_\epsilon \cap A_\epsilon),
\end{aligned}$$

so the result follows after one recalls that $\mathbb{P}(L_\epsilon, A_\epsilon) \geq \alpha$ and takes $\Gamma = \frac{1}{4}\min\{a_2 - a_3, a_1 - a_2\}\min\{\widehat{\alpha}, 1\}\alpha$. □

Appendix B. Proofs of Auxiliary Lemmas

Proof of Lemma 2. Because $\mathbb{E}[B] = pn$ and $(p + \varepsilon)n \leq k$, for any $u > 0$, we have the tail bound

$$\mathbb{P}((B - k)_+ \geq u) = \mathbb{P}(B - pn \geq k - pn + u) \leq \mathbb{P}(B - pn \geq \varepsilon n + u),$$

and Hoeffding's inequality (see, e.g., Boucheron et al. 2013, theorem 2.8) implies that

$$\mathbb{P}((B - k)_+ \geq u) \leq \exp\{-2\varepsilon^2 n - 4\varepsilon u - 2u^2/n\}, \qquad \text{for all } (p + \varepsilon)n \leq k.$$

By integrating both sides for $u \in [0, \infty)$, we then obtain, for all $(p + \varepsilon)n \leq k$, that

$$\mathbb{E}[(B - k)_+] \leq \int_0^\infty \mathbb{P}((B - k)_+ \geq u)\, du \leq \exp\{-2\varepsilon^2 n\}\int_0^\infty \exp\{-4\varepsilon u\}\, du \leq \frac{1}{4\varepsilon}.$$

To prove the second bound in (8), one applies the first bound to the binomial random variable $B' = n - B$ and the budget $k' = n - k$. □

Proof of Lemma 3. The index policy id $= \{\mathfrak{p}_{j,t} : j \in [m] \text{ and } t \in [n]\}$ defined by (33) is such that all values greater than or equal to $a_{j_{\mathrm{id}}-1}$ are selected together with a fraction of the $a_{j_{\mathrm{id}}}$ values. Formally, by Wald's lemma, we have that

$$\mathbb{E}[S_j^{\mathrm{id},\nu}] = \begin{cases} f_j \mathbb{E}[\nu] & \text{if } j \le j_{\mathrm{id}} - 1, \\ (k/n - \bar{F}(a_{j_{\mathrm{id}}}))\mathbb{E}[\nu] & \text{if } j = j_{\mathrm{id}}, \\ 0 & \text{if } j \ge j_{\mathrm{id}} + 1, \end{cases}$$

where $S_j^{\mathrm{id},\nu} = \sum_{t=1}^{\nu} \sigma_t^{\mathrm{id}} \mathbb{1}(X_t = a_j)$ is the number of a_j candidates selected by the index policy by time ν. Recall that now that for any policy π, we have the inequalities

$$V_{\mathrm{on}}^{\pi}(n,k) \le V_{\mathrm{off}}^{*}(n,k) \le DR(n,k) = \sum_{j=1}^{j_{\mathrm{id}}-1} a_j f_j n + a_{j_{\mathrm{id}}}(k - n\bar{F}(a_{j_{\mathrm{id}}})),$$

so

$$V_{\mathrm{off}}^{*}(n,k) - V_{\mathrm{on}}^{\mathrm{id}}(n,k) \le DR(n,k) - V_{\mathrm{on}}^{\mathrm{id}}(n,k) \le a_1 \mathbb{E}[n - \nu],$$

and to bound the regret as desired, it suffices to obtain an upper bound for $\mathbb{E}[n - \nu]$.

Let $\{B_t : t \in [n]\}$ be independent and identically distributed Bernoulli random variables with success probability $q = \sum_{i=1}^{j_{\mathrm{id}}-1} f_j + k/n - \bar{F}(a_{j_{\mathrm{id}}}) = k/n$. If

$$N_r = \sum_{t \in [r]} B_t - qr$$

is the centered number of candidates the index policy selects by time r, we then have, for $t \ge 0$ and $q = k/n$, that

$$n - \nu \ge t \quad \text{if and only if} \quad N_{n-\lceil t \rceil} \ge k - q(n - \lceil t \rceil) = q\lceil t \rceil.$$

In turn, Kolmogorov's maximal inequality (see, e.g., Billingsley 1995, theorem 22.4) tells us that for any $t > 0$,

$$\mathbb{P}\{n - \nu \ge t\sqrt{n}\} = \mathbb{P}\{N_{n-\lceil t\sqrt{n} \rceil} \ge q\lceil t\sqrt{n} \rceil\} \le \mathbb{P}(\sup_{t \in [n]} |N_t| \ge qt\sqrt{n}) \le \frac{\mathbb{E}[N_n^2]}{q^2 t^2 n} = \frac{1-q}{qt^2}.$$

It then follows that

$$\mathbb{E}\left[\frac{n-\nu}{\sqrt{n}}\right] \le 1 + \int_1^{\infty} \mathbb{P}(n - \nu \ge t\sqrt{n})\, dy \le \frac{1}{q},$$

so $\mathbb{E}[n - \nu] \le q^{-1}\sqrt{n} \le \varepsilon^{-1}\sqrt{n}$ for any $\varepsilon \in (0,1)$ and all pairs (n,k) such that $\varepsilon \le k/n$. □

Proof of Lemma 4. The nonadaptive policy $\widehat{\pi}$ that takes all values a_1 and rejects all others achieves bounded regret. To see this, notice that

$$\sum_{j=2}^{m} \mathfrak{S}_j^n = \sum_{j=2}^{m} \min\left\{Z_j^n, \left(k - \sum_{i \in [j-1]} Z_i^n\right)_+\right\} = (k - Z_1^n)_+.$$

Because the random variable Z_1^n is binomial with parameters n, f_1, and $0 \le k \le n(f_1 - \epsilon)$, Lemma 2 implies that

$$\mathbb{E}\left[\sum_{j=2}^{m} \mathfrak{S}_j^n\right] = \mathbb{E}[(k - Z_1^n)_+] \le \frac{1}{4\epsilon}.$$

In turn, the value of the offline solution for $k \le n(f_1 - \epsilon)$ satisfies the bound

$$V_{\mathrm{off}}^{*}(n,k) \le a_1 \mathbb{E}[\mathfrak{S}_1^n] + \frac{a_2}{4\epsilon} = a_1 \mathbb{E}[\min\{Z_1^n, k\}] + \frac{a_2}{4\epsilon}.$$

The nonadaptive policy $\widehat{\pi}$ takes all a_1 values and none of the others, until it runs out of budget at time $\nu = \min\{r \ge 1 : \sum_{t \in [r]} B_t \ge k \text{ or } r \ge n\}$, so $\mathbb{E}[S_j^{\widehat{\pi},n}] = 0$ for all $j \ge 2$ and

$$\mathbb{E}[S_1^{\widehat{\pi},n}] = \mathbb{E}\left[\sum_{t \in [\nu]} B_t \mathbb{1}(X_t = a_1)\right].$$

Furthermore, we have that $S_1^{\widehat{\pi},n} = Z_1^n$ if $Z_1^n < k$, and it equals k otherwise. Thus,

$$\mathbb{E}[S_1^{\widehat{\pi},n}] = \mathbb{E}[Z_1^{\nu}] = \mathbb{E}[\min\{Z_1^n, k\}] = \mathbb{E}[\mathfrak{S}_1^n],$$

so we finally have the bound

$$V^*_{\text{off}}(n,k) - \frac{a_2}{4\epsilon} \le a_1 \mathbb{E}[\min\{Z^n_1, k\}] = V^\pi_{\text{on}}(n,k),$$

just as needed. □

Proof of Lemma 5. Let $\mathcal{Z}$ denote a normal random variable with mean zero and variance one. Then

$$|\mathbb{E}[(N_n - \Upsilon\varsigma_n)_+] - \varsigma_n \mathbb{E}[(\mathcal{Z} - \Upsilon)_+]| \le d_W(N_n, \varsigma_n \mathcal{Z}) = \sup_{h \in \text{Lip}_1} |\mathbb{E}[h(N_n)] - \mathbb{E}[h(\varsigma_n \mathcal{Z})]|, \tag{B.1}$$

where Lip_1 is the set of Lipschitz-1 functions on $\mathbb{R}$. The random variables $B_1, B_2, \ldots, B_n$ are independent Bernoulli random variables with success probabilities $q_1, q_2, \ldots, q_n$ so that

$$\mathbb{E}[|B_t - q_t|^3] \le 2q_t(1-q_t) \qquad \text{and} \qquad \mathbb{E}[(B_t - q_t)^4] \le 2q_t(1-q_t)$$

and

$$\sum_{t \in [n]} \mathbb{E}[|B_t - q_t|^3] \le 2\varsigma_n^2 \qquad \text{and} \qquad \sum_{t \in [n]} \mathbb{E}[(B_t - q_t)^4] \le 2\varsigma_n^2.$$

A version of Stein's lemma (see, e.g., Ross 2011, theorem 3.6) for the sum of independent (not necessarily identically distributed) random variables tells us that

$$d_W(N_n/\varsigma_n, \mathcal{Z}) \le \frac{2}{\varsigma_n} + \frac{3\sqrt{2}}{\varsigma_n},$$

so, by the homogeneity of the distance function d_W, we also have that

$$d_W(N_n, \varsigma_n \mathcal{Z}) = \varsigma_n d_W(N_n/\varsigma_n, \mathcal{Z}) \le 2 + 3\sqrt{2}.$$

The inequality (B.1) then implies that

$$\varsigma_n \mathbb{E}[(\mathcal{Z} - \Upsilon)_+] - (2 + 3\sqrt{2}) \le \mathbb{E}[(N_n - \Upsilon\varsigma_n)_+],$$

and one can obtain an immediate lower bound for the left-hand side by

$$\varsigma_n \Upsilon \mathbb{P}(\mathcal{Z} \ge 2\Upsilon) \le \varsigma_n \mathbb{E}[(\mathcal{Z} - \Upsilon)\mathbb{1}(\mathcal{Z} \ge 2\Upsilon)] \le \varsigma_n \mathbb{E}[(\mathcal{Z} - \Upsilon)_+].$$

The left inequality in (34) then follows setting $\beta_1 = \Upsilon\mathbb{P}(\mathcal{Z} \ge 2\Upsilon)$. For the right inequality, we combine the earlier argument with the symmetry of the normal distribution and the Lipschitz-1 continuity of the map $x \mapsto (-x - \Upsilon\varsigma_n)_+$.

The argument for the second inequality in (35) is standard. Because $\mathbb{E}[N_n] = 0$ and $\mathbb{E}[N_n^2] = \varsigma_n^2$, we have that

$$\mathbb{E}\big[(N_n + \Upsilon\varsigma_n)_+^2\big] \le \mathbb{E}\big[(N_n + \Upsilon\varsigma_n)^2\big] = \varsigma_n^2 + \Upsilon^2\varsigma_n^2,$$

so taking $\beta_2 = 1 + \Upsilon^2$ concludes the proof. □

Proof of Lemma 6. Given a nonadaptive policy $\pi = \{p_{j,t} : j \in [m] \text{ and } t \in [n]\}$ and a constant $M < \infty$, we let

$$\bar{M} = \frac{M}{\min_{j \in [m-1]} |a_j - a_{j+1}|} \qquad \text{and} \qquad \iota = \arg\max_{j \in [m]} |s_j(\pi) - s_j^*|, \tag{B.2}$$

and we show that if

$$|s_\iota(\pi) - s_\iota^*| = \max_{j \in [m]} |s_j(\pi) - s_j^*| > \bar{M}\sqrt{n}, \tag{B.3}$$

then

$$DR(n,k) - V^\pi_{\text{on}}(n,k) > M\sqrt{n}.$$

We let $j_{\text{id}} \in [m]$ be the index such that $\bar{F}(a_{j_{\text{id}}}) \le k/n < \bar{F}(a_{j_{\text{id}}+1})$. There are two cases to consider: (i) (B.3) is attained at $\iota \ge j_{\text{id}} + 1$ ($k/n < \bar{F}(a_{j_{\text{id}}}) < \bar{F}(a_\iota)$), in which case $0 = s_\iota^* < s_\iota(\pi)$, and (ii) $\iota \le j_{\text{id}}$ ($\bar{F}(a_\iota \le \bar{F}(a_{j_{\text{id}}}) \le k/n$), in which case $0 < s_\iota^*$ and $s_\iota(\pi) < s_\iota^*$.

We begin with the first case; that is, we assume that (B.3) is attained for some $\iota \ge j_{\text{id}} + 1$ when $s_\iota^* = 0$, and we obtain that $s_\iota(\pi) \ge s_\iota^* + \bar{M}\sqrt{n} = \bar{M}\sqrt{n}$. To estimate the gap between $DR(n,k)$ and $V^\pi_{\text{on}}(n,k)$, consider a version of the deterministic relaxation

(15) that requires the selection of at least $\bar{M}\sqrt{n}$ candidates with ability a_ι out of the $f_\iota n$ available. Because $\bar{M} \le 2\epsilon\sqrt{n}$, we have that $\bar{M}\sqrt{n} \le f_j n$ for all $j \in [m]$, so we write the optimization problem as

$$\begin{aligned} DRC(n,k,\iota) = \max_{s_1,\ldots,s_m} & \sum_{j\in[m]} a_j s_j \\ \text{s.t.} \quad & 0 \le s_j \le \mathbb{E}[Z_j^n] \text{ for all } j \in [m], \\ & \bar{M}\sqrt{n} \le s_\iota, \\ & \sum_{j\in[m]} s_j \le k, \end{aligned}$$

and we obtain that

$$V_{\text{on}}^{\pi}(n,k) \le DRC(n,k,\iota) \le DR(n,k).$$

The unique maximizer $(\check{s}_1,\ldots,\check{s}_m)$ of $DRC(n,k,\iota)$ is given by

$$\check{s}_j = \begin{cases} s_j^* & \text{if } j \le j_{\text{id}} - 2, \\ s_{j_{\text{id}}-1}^* - (\bar{M}\sqrt{n} - s_{j_{\text{id}}}^*)_+ & \text{if } j = j_{\text{id}} - 1, \\ (s_{j_{\text{id}}}^*, - \bar{M}\sqrt{n})_+ & \text{if } j = j_{\text{id}}, \\ \bar{M}\sqrt{n} & \text{if } j = \iota, \\ 0 & \text{otherwise,} \end{cases}$$

so the difference between the value of the deterministic relaxation and the value of its constrained version is given by

$$DR(n,k) - DRC(n,k,\iota) = a_{j_{\text{id}}-1}(\bar{M}\sqrt{n} - s_{j_{\text{id}}}^*)_+ + a_{j_{\text{id}}}[s_{j_{\text{id}}}^* - (s_{j_{\text{id}}}^* - \bar{M}\sqrt{n})_+] - a_\iota \bar{M}\sqrt{n}.$$

Because $j_{\text{id}} - 1 < j_{\text{id}} \le \iota - 1 < \iota$, the monotonicity $a_\iota < a_{\iota-1} \le a_{j_{\text{id}}} < a_{j_{\text{id}}-1}$ gives us the lower bound

$$DR(n,k) - DRC(n,k,\iota) \ge a_{j_{\text{id}}-1}\bar{M}\sqrt{n} - a_\iota \bar{M}\sqrt{n} > [a_{\iota-1} - a_\iota]\bar{M}\sqrt{n},$$

and because $V_{on}^{\pi}(n,k) \le DRC(n,k,\iota)$, we have

$$DR(n,k) - V_{\text{on}}^{\pi}(n,k) > [a_{\iota-1} - a_\iota]\bar{M}\sqrt{n} \qquad \text{for all } \iota \ge j_{\text{id}} + 1. \tag{B.4}$$

A similar inequality can be obtained for second case in which (B.3) is attained at some index $\iota \le j_{\text{id}}$. In this case, we would consider a version of the deterministic relaxation (15) with the additional constraint $s_\iota \le nf_\iota - \bar{M}\sqrt{n}$ or $s_{j_{\text{id}}} \le k - n\bar{F}(a_{j_{\text{id}}}) - \bar{M}\sqrt{n}$. This analysis then implies the bound

$$DR(n,k) - V_{\text{on}}^{\pi}(n,k) > [a_\iota - a_{\iota+1}]\bar{M}\sqrt{n} \qquad \text{for all } \iota \le j_{\text{id}},$$

so if we recall (B.4) and use the definition of $\bar{M}$ in (B.2), we finally obtain that

$$DR(n,k) - V_{\text{on}}^{\pi}(n,k) > M\sqrt{n},$$

concluding the proof of the lemma. □

Proof of Lemma 7. Fix any nonadaptive policy $\pi \in \Pi_{\text{na}}$ and recall that $S_1^{\pi,\nu}$ is the number of a_1 candidates that the policy selects. If

$$\begin{aligned} \varphi_{-1}(z_2,\ldots,z_m,k) = \max_{s_2,\ldots,s_m} & \sum_{j=2}^{m} a_j s_j \\ \text{s.t.} \quad & 0 \le s_j \le z_j \quad \text{for all } j = 2,\ldots,m, \\ & \sum_{j=2}^{m} s_j \le k, \end{aligned}$$

then

$$V_{\text{on}}^{\pi}(n,k) \le a_1 \mathbb{E}[S_1^{\pi,\nu}] + \mathbb{E}[\varphi_{-1}(Z_2^n,\ldots,Z_m^n, k - S_1^{\pi,\nu})]. \tag{B.5}$$

Because $S_1^{\pi,\nu} \le \mathfrak{S}_1^n = \min\{Z_1^n, k\}$, the monotonicity in k of $\varphi_{-1}(\cdot,k)$ gives us the upper bound

$$\varphi_{-1}(Z_2^n,\ldots,Z_m^n, k - S_1^{\pi,\nu}) \le \varphi_{-1}(Z_2^n,\ldots,Z_m^n, k - \mathfrak{S}_1^n) + a_2(\mathfrak{S}_1^n - S_1^{\pi,\nu}),$$

so when we take expectations and recall (B.5), we obtain that

$$V_{\text{on}}^{\pi}(n,k) \le a_1 \mathbb{E}[S_1^{\pi,\nu}] + \mathbb{E}[\varphi_{-1}(Z_2^n,\ldots,Z_m^n, k - \mathfrak{S}_1^n)] + a_2(\mathbb{E}[\mathfrak{S}_1^n] - \mathbb{E}[S_1^{\pi,\nu}]).$$

We now note that $V^*_{\text{off}}(n,k) = a_1\mathbb{E}[\mathfrak{S}^n_1] + \mathbb{E}[\varphi_{-1}(Z^n_2,\ldots,Z^n_m,k-\mathfrak{S}^n_1)]$, so when we use this last decomposition in the displayed equation above, we conclude that

$$V^\pi_{\text{on}}(n,k) \le V^*_{\text{off}}(n,k) - (a_1-a_2)(\mathbb{E}[\mathfrak{S}^n_1] - \mathbb{E}[S^{\pi,\nu}_1]).$$

Let $M \equiv M(\epsilon,a_1,a_2)$ be such that the index policy achieves an $(a_1-a_2)M\sqrt{n}$ regret (see Lemma 3 with $\varepsilon = f_1+\epsilon$). Then, if π is such that $\mathbb{E}[S^{\pi,\nu}_1] < \mathbb{E}[\mathfrak{S}^n_1] - M\sqrt{n}$,

$$V^\pi_{\text{on}}(n,k) \le V^*_{\text{off}}(n,k) - (a_1-a_2)M\sqrt{n},$$

so π cannot be optimal. In other words, the optimal policy must satisfy

$$\mathbb{E}[S^{\pi,\nu}_1] \ge \mathbb{E}[\mathfrak{S}^n_1] - M\sqrt{n}. \tag{B.6}$$

Because $(f_1+\epsilon)n \le k$, we have from Lemma 2 that

$$\mathbb{E}[\mathfrak{S}^n_1] = \mathbb{E}[Z^n_1] - \mathbb{E}[(Z^n_1-k)_+] \ge nf_1 - \frac{1}{4\epsilon},$$

which, together with (B.6), implies that the optimal nonadaptive policy must have

$$\sum_{t\in[n]} p_{1,t}f_1 \ge \mathbb{E}[S^{\pi,\nu}_1] \ge nf_1 - \frac{1}{4\epsilon} - M\sqrt{n}.$$

Because $p_{1,t}f_1 \le q_t$ for all $t\in[n]$, it follows that there is another constant $M \equiv M(\epsilon,a_1,a_2)$ such that

$$\sum_{t\in[n]} \mathbb{1}\left\{q_t \le \frac{f_1}{2}\right\} \le \sum_{t\in[n]} \mathbb{1}\left\{p_{1,t} \le \frac{1}{2}\right\} \le M\sqrt{n},$$

concluding the proof of the left inequality of (36).

The argument for the right inequality of (36) is similar. We let

$$\begin{aligned}\varphi_{-m}(z_1,\ldots,z_{m-1},k) = \max_{s_1,\ldots,s_{m-1}} & \sum_{j=1}^{m-1} a_j s_j \\ \text{s.t.} \quad & 0 \le s_j \le z_j \text{ for all } j = 1,\ldots,m-1, \\ & \sum_{j=1}^{m-1} s_j \le k,\end{aligned}$$

and note that

$$V^\pi_{\text{on}}(n,k) \le a_m\mathbb{E}[S^{\pi,\nu}_m] + \mathbb{E}[\varphi_{-1}(Z^n_1,\ldots,Z^n_m,k-S^{\pi,\nu}_m)] \le a_m\mathbb{E}[S^{\pi,\nu}_m] + \mathbb{E}[\varphi_{-1}(Z^n_1,\ldots,Z^n_m,k)].$$

The decomposition $V^*_{\text{off}}(n,k) = \mathbb{E}[\varphi_{-m}(Z^n_1,\ldots,Z^n_{m-1},k)] + a_m\mathbb{E}[\mathfrak{S}^n_m]$ then implies that

$$V^\pi_{\text{on}}(n,k) \le V^*_{\text{off}}(n,k) - a_m(\mathbb{E}[\mathfrak{S}^n_m] - \mathbb{E}[S^{\pi,\nu}_m]).$$

Here we have that $k \le (1-f_m-\epsilon)n$, so Lemma 2 tells us that $\mathbb{E}[\mathfrak{S}^n_m] \le (4\epsilon)^{-1}$, and if $M \equiv M(\epsilon,a_m)$ is the constant such that the index policy achieves $a_mM\sqrt{n}$ regret, then we see that policy π cannot be optimal if $\mathbb{E}[S^{\pi,\nu}_m] > M\sqrt{n}$. Because $(1-p_{m,t})f_m \le 1-q_t$, this observation then gives us another constant $M \equiv M(\epsilon,a_m)$ such that

$$\sum_{t\in[n]} \mathbb{1}\left\{1-q_t \le \frac{f_m}{2}\right\} \le \sum_{t\in[n]} \mathbb{1}\left\{1-p_{m,t} \le \frac{1}{2}\right\} = \sum_{t\in[n]} \mathbb{1}\left\{p_{m,t} \ge \frac{1}{2}\right\} \le M\sqrt{n},$$

which completes the proof. □

Appendix C. The Dynamic-Programming Formulation

Let $v_\ell(w,\kappa)$ denote the expected total ability when ℓ candidates are yet to be inspected (the are ℓ periods to go), the current level of accrued ability is w, and at most κ can be selected (κ is the residual recruiting budget). The sequence of functions $\{v_\ell : \mathbb{R}_+\times\mathbb{Z}_+ \to \mathbb{R}_+ \text{ for } 1 \le \ell < \infty\}$ then satisfies the Bellman recursion

$$v_\ell(w,\kappa) = \sum_{j\in[m]} \max\{v_{\ell-1}(w+a_j,\kappa-1), v_{\ell-1}(w,\kappa)\}f_j, \tag{C.1}$$

with the boundary conditions

$$v_0(w,\kappa) = w \text{ for all } \kappa \in \mathbb{Z}_+ \text{ and } w \geq 0 \qquad \text{and} \qquad v_\ell(w,0) = w \text{ for all } \ell \in [n] \text{ and } w \geq 0.$$

If n is the number of available candidates and k is the initial recruiting budget, the optimality principle of dynamic programming then implies that

$$V^*_{\text{on}}(n,k) = v_n(0,k) \qquad \text{for all } n,k \in \mathbb{Z}_+.$$

The Bellman equation (C.1) implies that the optimal online policy takes the form of a threshold policy, and the next proposition proves that the optimal thresholds depend only on the residual budget κ and not on the accrued ability w.

Proposition 5 (State-Space Reduction). *Let $\{g_\ell : \mathbb{Z}_+ \to \mathbb{R}_+ \text{ for } 1 \leq \ell < \infty\}$ satisfy the recursion*

$$g_\ell(\kappa) = \sum_{j\in[m]} \max\{a_j + g_{\ell-1}(\kappa-1), g_{\ell-1}(\kappa)\} f_j, \tag{C.2}$$

with the boundary conditions

$$g_0(\kappa) = 0 \text{ for al } \kappa \in \mathbb{Z}_+ \quad \text{and} \quad g_\ell(0) = 0 \text{ for all } 1 \leq \ell < \infty. \tag{C.3}$$

Then

$$v_\ell(w,\kappa) = w + g_\ell(\kappa) \qquad \text{for all } w \in \mathbb{R}_+ \text{ and } \kappa \in \mathbb{Z}_+. \tag{C.4}$$

Consequently, with ℓ periods to go and a residual budget of κ, it is optimal to select a candidate with ability a_j if and only if

$$a_j \geq h_\ell(\kappa) \equiv g_{\ell-1}(\kappa) - g_{\ell-1}(\kappa-1).$$

Under the optimal online policy, ability levels strictly smaller than $h_\ell(\kappa)$ are skipped and the rest are selected. Because the Markov decision problem associated with the Bellman equation (C.2) is a finite-horizon problem with finite state space and a finite number of actions, this deterministic policy is also optimal within the larger class of nonanticipating policies (see, e.g., Bertsekas and Shreve 1978, corollary 8.5.1).

Proof of Proposition 5. This is an induction proof. For $\ell = 1$, we have by (C.1) that

$$v_1(w,0) = w \quad \text{and} \quad v_1(w,\kappa) = w + \sum_{j\in[m]} a_j f_j = w + \mathbb{E}[X_1] \text{ for all } \kappa \geq 1.$$

Taking

$$g_1(\kappa) = \begin{cases} 0 & \text{if } \kappa = 0, \\ \mathbb{E}[X_1] & \text{if } \kappa \geq 1, \end{cases}$$

one has that

$$v_1(w,\kappa) = w + g_1(\kappa) \qquad \text{for all } w \geq 0 \text{ and all } \kappa \in \mathbb{Z}_+.$$

Next, as the induction hypothesis, suppose that we have the decomposition

$$v_{\ell-1}(w,\kappa) = w + g_{\ell-1}(\kappa) \qquad \text{for all } w \geq 0 \text{ and all } \kappa \in \mathbb{Z}_+.$$

If $\kappa = 0$, the decomposition is satisfied with $g_\ell(0) = 0$ and

$$v_\ell(w,0) = w + g_\ell(0) = w. \tag{C.5}$$

For $\kappa \geq 1$, the Bellman equation (C.1) and the induction assumption imply that

$$\begin{aligned} v_\ell(w,\kappa) &= w + \sum_{j\in[m]} \max\{a_j + v_{\ell-1}(w+a_j, \kappa-1) - w - a_j, v_{\ell-1}(w,\kappa) - w\} f_j \\ &= w + \sum_{j\in[m]} \max\{a_j + g_{\ell-1}(\kappa-1), g_{\ell-1}(\kappa)\} f_j. \end{aligned}$$

Defining recursively

$$g_\ell(\kappa) = \sum_{j\in[m]} \max\{a_j + g_{\ell-1}(\kappa-1), g_{\ell-1}(\kappa)\} f_j \qquad \text{for all } \kappa \geq 1,$$

we then have, together with (C.5), that (C.4) holds for all $w \geq 0$ and all $\kappa \in \mathbb{Z}_+$. From this argument, it also follows that $g_\ell(\kappa)$ satisfies the recursion (C.2) with the boundary condition (C.3). □

Appendix D. Deterministic Relaxation Revisited

The subset $\mathcal{T}'$ is identified by taking $0<\epsilon'<\epsilon=\frac{1}{2}\min\{f_m,f_{m-1},\ldots,f_1\}$ and letting

$$\mathcal{T}'=\{(n,k)\in\mathcal{T}:\bar{F}(a_j)+\epsilon'\le k/n\le\bar{F}(a_{j+1})-\epsilon' \text{ for some } j\in[m]\}.$$

Proposition 6 (Deterministic-Relaxation Gap). *There exists a constant $M\equiv M(\epsilon,m,a_m,a_{m-1},\ldots,a_1)$ such that*

$$0\le DR(n,k)-V^*_{\text{off}}(n,k)\le M\sqrt{n}\qquad \text{for } (n,k)\in\mathcal{T}.$$

Furthermore,

$$0\le DR(n,k)-V^*_{\text{off}}(n,k)\le\frac{a_1 m}{4\epsilon'}\qquad \text{for } (n,k)\in\mathcal{T}'.$$

The proof of Proposition 6 requires the following auxiliary lemma.

Lemma 8. *There exists a constant $M\equiv M(\epsilon,m,a_m,a_{m-1},\ldots,a_1)$ such that, for all $j\in[m]$,*

$$\mathbb{E}[\mathfrak{S}^n_j]=\mathbb{E}\left[\min\left\{Z^n_j,\left(k-\sum_{i\in[j-1]}Z^n_i\right)_+\right\}\right]=\min\left\{\mathbb{E}[Z^n_j],\left(k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]\right)_+\right\}\pm M\sqrt{n}.$$

In turn,

$$\mathbb{E}[\mathfrak{S}^n_j]=s^*_j\pm M\sqrt{n},\ \text{for all } j\in[m].$$

Proof. Let $\xi^n_j=Z^n_j-\mathbb{E}[Z^n_j]$ for all $j\in[m]$ and obtain the representation

$$\min\left\{Z^n_j,\left(k-\sum_{i\in[j-1]}Z^n_i\right)_+\right\}=\min\left\{\mathbb{E}[Z^n_j]+\xi^n_j,\left(k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]-\sum_{i\in[j-1]}\xi^n_i\right)_+\right\}.$$

For any real numbers a,b,c, and d, we have that

$$\min\{a,\max\{c,0\}\}-|b|-|d|\le\min\{a+b,\max\{c+d,0\}\}\le\min\{a,\max\{c,0\}\}+|b|+|d|,$$

so by setting $a=\mathbb{E}[Z^n_j]$, $b=\xi^n_j$, $c=k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]$, and $d=-\sum_{i\in[j-1]}\xi^n_i$ and taking expectations, we have

$$\mathbb{E}\left[\min\left\{Z^n_j,\left(k-\sum_{i\in[j-1]}Z^n_i\right)_+\right\}\right]=\mathbb{E}\left[\min\left\{\mathbb{E}[Z^n_j],\left(k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]\right)_+\right\}\right]\pm\sum_{i\in[j]}\mathbb{E}[|\xi^n_i|].$$

For the random variables $Z^n_1,\ldots,Z^n_m$, we have $\text{Var}[Z^n_j]=\mathbb{E}[(\xi^n_j)^2]=nf_j(1-f_j)$ for all $j\in[m]$, and by the Cauchy–Schwartz inequality,

$$\mathbb{E}\left[\min\left\{Z^n_j,\left(k-\sum_{i\in[j-1]}Z^n_i\right)_+\right\}\right]=\mathbb{E}\left[\min\left\{\mathbb{E}[Z^n_j],\left(k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i]\right)_+\right\}\right]\pm\sum_{i\in[j]}\sqrt{nf_i(1-f_i)},$$

and the result follows by setting $M=\sum_{j\in[m]}\sqrt{f_j(1-f_j)}$. □

Proof of Proposition 6. The first part follows immediately from Lemma 8. Next, for $(n,k)\in\mathcal{T}'$ and the index j_{id} such that $\bar{F}(a_{j_{\text{id}}})\le k/n<\bar{F}(a_{j_{\text{id}}+1})$, the optimal solution (16) of the deterministic relaxation $DR(n,k)$ is given by

$$s^*_j=\begin{cases}\mathbb{E}[Z^n_j] & \text{if } j\le j_{\text{id}}-1,\\ k-\sum_{i\in[j-1]}\mathbb{E}[Z^n_i] & \text{if } j=j_{\text{id}},\\ 0 & \text{otherwise.}\end{cases}\tag{D.1}$$

To estimate the deterministic-relaxation gap when $(n,k)\in\mathcal{T}'$, it suffices to study the differences $s^*_j-\mathbb{E}[\mathfrak{S}^n_j]=\mathbb{E}[Z^n_j-\mathfrak{S}^n_j]$ for $j\le j_{\text{id}}-1$ and $s^*_{j_{\text{id}}}-\mathbb{E}[\mathfrak{S}^n_{j_{\text{id}}}]=\mathbb{E}[k-\sum_{i\in[j_{\text{id}}-1]}Z^n_i-\mathfrak{S}^n_{j_{\text{id}}}]$. Recalling the definition of $\mathfrak{S}^n_j$ in (3), we have

$$0\le Z^n_j-\mathfrak{S}^n_j\le\begin{cases}0 & \text{if } k-\sum_{i\in[j-1]}Z^n_i\ge Z^n_j,\\ \sum_{i\in[j]}Z^n_i-k & \text{if } k-\sum_{i\in[j-1]}Z^n_i<Z^n_j.\end{cases}$$

In particular,

$$0 \le Z_j^n - \min\left\{Z_j^n, \left(k - \sum_{i\in[j-1]} Z_i^n\right)_+\right\} \le \left(\sum_{i\in[j]} Z_i^n - k\right)_+.$$

Taking expectations on both sides, and recalling (D.1) for $j \le j_{\mathrm{id}} - 1$, we have

$$0 \le s_j^* - \mathbb{E}[\mathfrak{S}_j^n] \le \mathbb{E}\left[\left(\sum_{i\in[j]} Z_i^n - k\right)_+\right].$$

For such $j \le j_{\mathrm{id}} - 1$, the sum $\sum_{i\in[j]} Z_i^n$ is a binomial random variable with n trials and success probability $\bar{F}(a_{j+1}) \le \bar{F}(a_{j_{\mathrm{id}}})$, so, because $(n,k) \in \mathcal{T}'$ and $(\bar{F}(a_{j_{\mathrm{id}}}) + \epsilon')n \le k$, we obtain from (8) that

$$s_j^* = \mathbb{E}[\mathfrak{S}_j^n] \pm \frac{1}{4\epsilon'}, \qquad \text{for all } j \le j_{\mathrm{id}} - 1. \tag{D.2}$$

Similarly,

$$\left\{k - \sum_{i\in[j_{\mathrm{id}}-1]} Z_i^n\right\} - \mathfrak{S}_{j_{\mathrm{id}}}^n = \max\left\{k - \sum_{i\in[j_{\mathrm{id}}]} Z_i^n, \min\left\{0, k - \sum_{i\in[j_{\mathrm{id}}-1]} Z_i^n\right\}\right\},$$

so we have the two inequalities

$$\min\left\{0, k - \sum_{i\in[j_{\mathrm{id}}-1]} Z_i^n\right\} \le \left\{k - \sum_{i\in[j_{\mathrm{id}}-1]} Z_i^n\right\} - \mathfrak{S}_{j_{\mathrm{id}}}^n \le \max\left\{k - \sum_{i\in[j_{\mathrm{id}}]} Z_i^n, 0\right\}.$$

Here the two sums $\sum_{i\in[j_{\mathrm{id}}-1]} Z_i^n$ and $\sum_{i\in[j_{\mathrm{id}}]} Z_i^n$ are again binomial random variables with n trials and success probabilities given, respectively, by $\bar{F}(a_{j_{\mathrm{id}}})$ and $\bar{F}(a_{j_{\mathrm{id}}+1})$. Taking expectations and using $\bar{F}(a_{j_{\mathrm{id}}}) + \epsilon' \le k/n \le \bar{F}(a_{j_{\mathrm{id}}+1}) - \epsilon'$, Lemma 2 guarantees that

$$-\frac{1}{4\epsilon'} \le k - \sum_{i\in[j_{\mathrm{id}}-1]} \mathbb{E}[Z_i^n] - \mathbb{E}[\mathfrak{S}_{j_{\mathrm{id}}}^n] \le \frac{1}{4\epsilon'}.$$

In turn, the representation (D.1) for $s_{j_{\mathrm{id}}}^*$ implies that

$$s_{j_{\mathrm{id}}}^* = \mathbb{E}[\mathfrak{S}_{j_{\mathrm{id}}}^n] \pm \frac{1}{4\epsilon'}. \tag{D.3}$$

Combining the estimates (D.2) and (D.3) then gives us that

$$DR(n,k) - V_{\mathrm{off}}^*(n,k) \le \frac{a_1 m}{4\epsilon'} \qquad \text{for all } (n,k) \in \mathcal{T}',$$

just as needed to complete the proof of the proposition. □

References

Babaioff M, Immorlica N, Kempe D, Kleinberg R (2007) A knapsack secretary problem with applications. Goemans M, Jansen K, Rolim JDP, Trevisan L, eds. *Approximation, Randomization Combinatorial Optimization: Algorithms and Techniques*, Lecture Notes in Computer Science, vol. 2129 (Springer-Verlag, Berlin, Heidelberg), 16–28.

Bertsekas D, Shreve S (1978) *Stochastic Optimal Control: The Discrete Time Case* (Academic Press, New York).

Billingsley P (1995) *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 3rd ed. (John Wiley & Sons, Inc., New York).

Boucheron S, Lugosi G, Massart P (2013) *Concentration Inequalities* (Oxford University Press, Oxford, UK).

Bramson M, et al. (2008) Stability of queueing networks. *Probab. Surveys* 5:169–345.

Cayley A (1875) Mathematical questions with their solutions. *Ed. Times* 23:18–19.

Chow YS, Moriguti S, Robbins H, Samuels SM (1964) Optimal selection based on relative rank (the "secretary problem"). *Israel J. Math.* 2(2):81–90.

Ferguson TS (1989) Who solved the secretary problem? *Statist. Sci.* 4(3):282–296.

Freeman PR (1983) The secretary problem and its extensions: A review. *Internat. Statist. Rev.* 51(2):189–206.

Gilbert JP, Mosteller F (1966) Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* 61(313):35–73.

Hajek B (1982) Hitting-time and occupation-time bounds implied by drift analysis with applications. *Adv. Appl. Probab.* 14(3):502–525.

Kleinberg R (2005) A multiple-choice secretary algorithm with applications to online auctions. *Proc. 16th Annual ACM-SIAM Sympos. Discrete Algorithms* (ACM, New York), 630–631.

Kleinberg R (2017) Personal communication with the authors on the regret for the multi-secretary problem with (n,k)-dependent distribution, September 26, 2017.

Kleywegt AJ, Papastavrou JD (1998) The dynamic and stochastic knapsack problem. *Oper. Res.* 46(1):17–35.

Lindley DV (1961) Dynamic programming and decision theory. *Appl. Statist.* 10(1):39–51.

Louchard G, Bruss FT (2016) Sequential selection of the κ best out of n rankable objects. *Discrete Math. Theor. Comput. Sci.* 18(3):13.
Moser L (1956) On a problem of Cayley. *Scripta Mathematica* 22(3/4):289–292.
Ross N (2011) Fundamentals of Stein's method. *Probab. Surveys* 8:210–293.
Talluri KT, van Ryzin GJ (2004) *The Theory and Practice of Revenue Management*, International Series in Operations Research & Management Science, vol. 68 (Kluwer Academic Publishers, Boston).
Wu H, Srikant R, Liu X, Jiang C (2015) Algorithms with logarithmic or sublinear regret for constrained contextual bandits. Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, eds. *Advances in Neural Information Processing Systems* (Curran Associates, Red Hook, New York), 433–441.