THE SECRETARY PROBLEM WITH A STOCHASTIC SIGNAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning-augmented approaches to the secretary problem typically use value estimates, rank predictions, or explicit recommendations. We study advice conveyed solely through timing: a content-free stochastic signal that arrives no later than the best of n items. Under uniformly random arrival order, we characterize the exact finite-n optimal policy and success probability for arbitrary known signal distributions, assuming independence of the signal and relative order conditional on the best item's position. For the α-power family, where increasing α shifts the signal towards the best item's arrival, we derive optimal threshold policies and closed-form asymptotics. Specifically, the limiting success probability exceeds the classic $1/e$ benchmark for every $\alpha > 0$, equals $1/2$ for the uniform case $\alpha = 1$, and approaches 1 as $\alpha \to \infty$. We quantify robustness to misspecifying α by showing that conservative estimates preserve the classic $1/e$ benchmark asymptotically. Under adversarial arrival order, we determine the exact worst-case guarantees for the α-power family, including constant success probabilities when $\alpha = \Theta(n)$.

1 INTRODUCTION

The classic secretary problem is one of the most fundamental models of online decision making under uncertainty (Dynkin, 1963; Lindley, 1961; Gilbert & Mosteller, 1966; Ferguson, 1989). Its parsimonious formulation has made it a natural testbed for richer information models. Recent work on secretary and related problems studies algorithms with access to samples (Kaplan et al., 2025; Correa et al., 2025), augmented by predictions of values, ranks, or thresholds (Antoniadis et al., 2023b; Fujii & Yoshida, 2024; Braun & Sarkar, 2024; Balkanski et al., 2024; Nourmohammadi et al., 2026; Karisani et al., 2026), deciding when to query a perfect or noisy expert subject to a query budget (Liu et al., 2023; Benomar & Perchet, 2023; Moustakides et al., 2024), or relying on more general forms of advice (Dütting et al., 2024). In many of these models, auxiliary information is supplied *before* the decision making starts or *synchronous* with the item it describes, and contains an explicit value, rank, or recommendation; we give an overview in Appendix A.

In many applications, however, side information is generated by a separate process and arrives on a separate, *asynchronous* timeline. As an example, a hiring platform may be unable to disclose applicant identities or scores, but an anonymous alert generated before or when its highest-ranked applicant is presented still rules out all earlier applicants as the highest-ranked one. Related temporal indicators are used in popularity forecasting through early attention measurements (Szabó & Huberman, 2010; Tatar et al., 2014; Hu et al., 2017), in recommendation systems through evolving user-behavior signals (Yang & Zhan, 2022; Joulani et al., 2013; Qian et al., 2024), and in scientific impact prediction or scouting through early indicators of future long-run success (Dziallas & Blind, 2019; Abramo et al., 2019; Adams, 2005). Such indicators may contain rich information in practice. Our model isolates only their *timing* relative to a target event.

This observation motivates a question at the boundary of optimal stopping theory and learning-augmented online algorithms: *How valuable is the temporal information conveyed by a content-free stochastic signal?* We focus on the simplest reliable case: the signal arrives no later than the best item, and its arrival time (together with its absence beforehand) is the entire additional observation.

In the standard secretary problem, n valued items, indexed by $\{1, \dots, n\} =: [n]$, arrive online one by one, in either adversarial or uniformly random order, at times $1, \dots, n$. The goal is to *maximize the probability* of stopping exactly when the unique best-valued item arrives. Without

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

side information, the optimal random-order success probability converges to $1/e$ (Dynkin, 1963; Lindley, 1961; Ferguson, 1989). Adversarial order is much harsher: deterministic algorithms cannot guarantee positive success probability and stopping uniformly at random gives $1/n$, which is optimal.

We introduce a stochastic signal model for secretary problems. After the arrival order and in particular the arrival time I of the unique best item has been fixed, a time S is drawn from a distribution supported on $[I]$, which only depends on I. At every time, the algorithm observes the current item and its relative rank among the items seen so far. At time S, before deciding whether to stop, the algorithm observes the signal's arrival. Since $S \leq I$, every item before S can be ruled out, but the signal neither identifies nor evaluates the best item. Revealing S before the process starts would not change the optimal value because stopping before S cannot succeed. Our results therefore concern the information conveyed by the signal's timing. Our main question is how the conditional distribution of S given I determines the optimal policy and value.

The main message is that this question admits an exact answer. We give a high-level overview now and precise statements and definitions in Section 1.1:

- **Random-order full characterization.** For every signal distribution, we give an explicit finite-n recursion for the optimal policy and value.
- **Random-order closed forms for α-power signals.** In random order, we obtain exact optimal threshold policies for every $\alpha > 0$. In particular, the asymptotic optimum is always above $1/e$, and approaches 1 as α grows. We also quantify robustness to misspecified α and give finite-sample policy guarantees for learning α from historical instances.
- **Adversarial-order characterization for α-power signals.** We determine the exact optimal deterministic and randomized guarantees for α-power signals for every $\alpha > 0$. Specifically, constant success is possible if α scales linearly with n.

Conceptually, our results show that *timing alone* can be algorithmically valuable. The signal cannot be used to assess the quality of the current item directly. Nevertheless, in the α-power family it beats the classic random-order benchmark for every $\alpha > 0$, and signals whose concentration grows as $\alpha = \Theta(n)$ yield constant guarantees even under adversarial order.

1.1 Model and detailed results

We now formally introduce our models. We are given an arbitrary conditional signal distribution $q_i(s) := \mathbb{P}[S = s \mid I = i]$ supported on $s \leq i$ for each $i \in [n]$. In random order, conditional on I, the signal is independent of the relative order of the items. For closed-form guarantees and throughout the adversarial-order analysis we focus on the *α-power family* for $\alpha > 0$. Conditioned on $I = i$, the signal time is distributed as $S = \lceil i \cdot B \rceil$ for $B \sim \mathrm{Beta}(\alpha, 1)$; equivalently, each signal time $s \in [i]$ has probability $\mathbb{P}[S = s \mid I = i] = \frac{s^\alpha - (s-1)^\alpha}{i^\alpha}$. This family of distributions gives a simple interpolation between signals that tend to occur early ($\alpha \ll 1$) and signals that tend to occur late ($\alpha \gg 1$) in the range $[i]$. Two prominent special cases are $\alpha = 1$, where S is uniform on $[i]$, and integer values $\alpha = m \geq 1$, where S has the same distribution as the maximum of m independent uniform signals in $[i]$. We visualize representative densities in Figure 1.

We call time t a *record time* and the item arriving at t a *record* if it is strictly better than every preceding item. For any integer-valued possibly random threshold $0 \leq T \leq n$ generated from observed signals and independent internal randomness, the threshold policy $\mathcal{A}(T)$ rejects all items before time T and then accepts the first record at or after T.

Random order: general signal distributions (Section 2). For every signal distribution, we characterize the exact optimal policy and value. Both are computable by dynamic programming in $O(n^2)$ time. Interestingly, for some signal distributions, no threshold policy is optimal. We give a necessary and sufficient condition for stopping at a post-signal record to be optimal, which in particular implies that $\mathcal{A}(S)$ is optimal. We also specialize the criterion to a broad family of separable distributions and express the value of $\mathcal{A}(S)$ in terms of the conditional means $\mathbb{E}[S \mid I = i]$.

Closed forms for the α-power family (Section 2). For the α-power family, we show exact finite-n threshold policies and closed-form asymptotics (Theorem 2). If $\alpha \geq 1$, the optimal rule is $\mathcal{A}(S)$, which does not need to know n. If $0 < \alpha < 1$, the optimal policy is $\mathcal{A}(\max\{S, k_n\})$, where

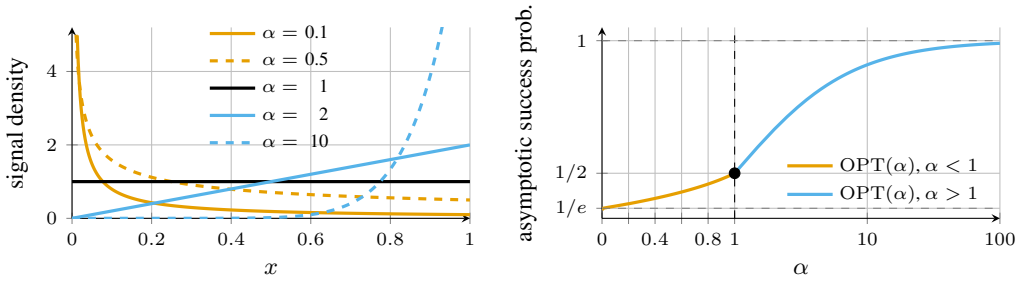

Figure 1: Left: PDFs of $\mathrm{Beta}(\alpha, 1)$. Right: Plot of $\mathrm{OPT}(\alpha)$ from Theorem 2.

$\frac{k_n}{n} \to (1-\alpha)^{1/\alpha}$. The limiting optimal value is

$$\mathrm{OPT}_n(\alpha) \xrightarrow{n\to\infty} \mathrm{OPT}(\alpha) = \begin{cases} \dfrac{\alpha + (1-\alpha)^{1+1/\alpha}}{\alpha+1}, & \text{if } 0 < \alpha < 1, \\ \dfrac{\alpha}{\alpha+1}, & \text{if } \alpha \geq 1. \end{cases}$$

These policy characterizations hold for every finite n. In particular, $\mathrm{OPT}_n(1) = \frac{n+1}{2n}$, and, for every fixed $\alpha \geq 1$, we have $\mathrm{OPT}_n(\alpha) = \frac{\alpha}{\alpha+1} + O_\alpha(\frac{\log n}{n})$. Thus, $\mathrm{OPT}(\alpha) > 1/e$ for every $\alpha > 0$, while $\mathrm{OPT}(\alpha) \to 1/e$ as $\alpha \to 0$, because the relative signal time S/I concentrates near zero.

Robustness for the α-power family (Section 3). If $\alpha \geq 1$, then $\mathcal{A}(S)$ is optimal even without knowing α. More generally, if the policy uses a conservative estimate $\hat{\alpha} \leq \alpha$, we still achieve at least $\mathrm{OPT}(\hat{\alpha}) > 1/e$ asymptotically (Theorem 3). We derive the exact limiting success probability for every pair $(\alpha, \hat{\alpha})$. Sufficiently large overestimates can reduce the success probability below $1/e$.

Adversarial order: deterministic and randomized policies (Section 4). When the position $i^\star$ of the best item is adversarially fixed, the exact randomized optimum is $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha) = \frac{n^\alpha}{\sum_{j=1}^n j^\alpha}$ and attained by $\mathcal{A}(\max\{S, R\})$ for a suitable random threshold R (Theorem 4). The deterministic optimum is $\mathrm{OPT}_n^{\mathrm{det}}(\alpha) = 1 - (1-1/n)^\alpha$ and attained by $\mathcal{A}(S)$ (Theorem 6). For every positive sequence (α_n), both values are $o(1)$ if $\alpha_n = o(n)$ and converge to $1 - e^{-c}$ if $\alpha_n/n \to c \in (0, \infty)$ (Corollaries 5 and 7). Thus, randomization improves the finite-n guarantee, but this advantage vanishes asymptotically when $\alpha_n/n \to c \in (0, \infty)$.

Full signal histories (Section 5). For integer $\alpha = m$, an α-power signal can be viewed as the last of m independent uniform signals in $[i^\star]$. Observing *all* m signal times does not improve the optimal randomized worst-case success probability (Theorem 8), but can improve the deterministic guarantee. We characterize the latter by an integer linear program (Theorem 9). For $m = 2$, the improvement is already strict at $n = 4$ and approaches $3/2$ asymptotically (Theorem 18).

Both arrival orders: experimental results (Section 6). We complement the theory with simulations of the actual online policies. In the random-order model, we show that the theoretically predicted gains over the classic secretary benchmark are visible already at moderate problem sizes, and that the threshold policies behave robustly when the signal-quality parameter is underestimated. We further test corrupted signals and show that a simple fallback mitigates their effect, although it need not retain the classic benchmark. Our experiments compare fitted policies with a baseline requiring no parameter estimate and test distribution-family misspecification and rank dependence. Learning curves compare the policies obtained by conservative estimation (Theorem 16) and maximum-likelihood estimation (MLE) as the number of training episodes grows. In the adversarial-order model, we confirm the behavior in the $\alpha = cn$ regime predicted by the analysis, and demonstrate that full signal histories can strictly help deterministic algorithms.

Further related work and relation to secretary with advice. Our random-order model is a special case of the joint rank/signal framework of Dütting et al. (2024), whose LP and structural

results apply. Our contribution within that framework is the explicit $O(n^2)$ recursion, a signal-specific threshold analysis and closed forms of optimal success probabilities. Moreover, we also give adversarial-order characterizations. We give further details on this relation in Appendix A.

2 RANDOM-ORDER MODEL: OPTIMAL POLICIES

We start with the random-order model: items have distinct values and arrive in uniformly random order. The policy observes relative ranks and the signal history. Here, the arrival time I of the best item is uniformly distributed over $[n]$, that is, $\mathbb{P}[I = i] = 1/n$ for each $i \in [n]$. The classic benchmark uses an integer threshold near n/e; it is asymptotically optimal, and its success probability converges to $1/e \approx 0.37$ (Dynkin, 1963; Lindley, 1961; Ferguson, 1989).

2.1 THE α-POWER FAMILY: WARM-UP

Single uniform signal. We show that we can improve substantially over $1/e$ even with a single conditionally uniform signal $[S \mid I = i] \sim \mathcal{U}([i])$ (equivalent to an α-power signal with $\alpha = 1$).

We consider the policy $\mathcal{A}(S)$, that is, we reject all items before time S and accept the first record in $\{S, \dots, n\}$. For the analysis, fix a realization $I = i \in [n]$ and $S = s \in [i]$. If $i = s$, then $\mathcal{A}(S)$ picks i and succeeds. If $s < i$ (and hence $i \geq 2$), then $\mathcal{A}(S)$ wins if the best item among $[i-1]$ arrives before s. This event has probability $\frac{s-1}{i-1}$. When it occurs, no record appears from time s through $i-1$, so $\mathcal{A}(S)$ stops at i. For $i \geq 2$, averaging over $s \in [i]$ yields success probability $\frac{1}{i}\sum_{s=1}^{i} \frac{s-1}{i-1} = \frac{1}{2}$. For $i = 1$, the policy succeeds with probability 1. Averaging over $i \in [n]$ gives an overall success probability of $\frac{1}{n} + \frac{n-1}{2n} = \frac{n+1}{2n} > \frac{1}{2}$.

α-Power signals. As a second warm-up, we extend the preceding argument to an α-power signal with an arbitrary fixed parameter $\alpha > 0$. We again use the policy $\mathcal{A}(S)$: reject all items before the signal time S and accept the first record from time S onward.

For the analysis, condition on $I = i$. The case $i = 1$ is trivial as then $S = 1$ and the policy succeeds. Let $i \geq 2$ and condition on $S = s \in [i]$. If $s = i$, then the signal occurs at the best item and $\mathcal{A}(S)$ accepts it. If $s < i$, then $\mathcal{A}(S)$ succeeds exactly when the best item among the first $i-1$ items appears before time s, which happens with probability $\frac{s-1}{i-1}$. Therefore, summation by parts gives

$$\mathbb{P}[\text{success} \mid I = i] = 1 - \frac{1}{(i-1)i^\alpha} \sum_{s=1}^{i-1} s^\alpha = \frac{\alpha}{\alpha+1} + O_\alpha\left(\frac{1}{i}\right). \tag{1}$$

Averaging over $i \in [n]$ yields

$$\mathbb{P}[\text{success}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}[\text{success} \mid I = i] = \frac{\alpha}{\alpha+1} + O_\alpha\left(\frac{\log n}{n}\right) \xrightarrow{n\to\infty} \frac{\alpha}{\alpha+1}. \tag{2}$$

The limiting success probability is increasing in α and tends to 1 as $\alpha \to \infty$, so increasingly late signals make $\mathcal{A}(S)$ nearly perfect. In Section 2.3, we show that $\mathcal{A}(S)$ is indeed optimal for all $\alpha \geq 1$. As $\alpha \to 0$, the asymptotic success probability of $\mathcal{A}(S)$ tends to zero. In this regime, the optimal policy combines S with a deterministic threshold. Its asymptotic success probability remains strictly above $1/e$ for every $0 < \alpha < 1$ and converges to $1/e$ as $\alpha \to 0$.

2.2 ARBITRARY SIGNAL DISTRIBUTIONS

We now allow an arbitrary conditional signal distribution $q_i(s) := \mathbb{P}[S = s \mid I = i]$ supported on $s \leq i$. Conditional on I, the signal is independent of the relative order of the items. The policy knows n and $(q_i(s))_{s \leq i}$. For $s \leq t$, let $\Psi_{t,s} := tq_t(s) + \sum_{i=t+1}^{n} q_i(s)$ be the posterior normalizing term. We call a tuple (t, s) of time t and signal time s *feasible* if $\Psi_{t,s} > 0$.

Theorem 1. *For each $s \in [n]$, set $F_{n+1,s} := 0$ and, for $t = n, n-1, \dots, s$, define*

$$F_{t,s} := \max\left\{F_{t+1,s}, q_t(s) + \left(1 - \frac{1}{t}\right) F_{t+1,s}\right\}. \tag{3}$$

There is an optimal deterministic policy that never stops before the signal and only stops at records. At every feasible record state (t, s) with observed signal $S = s$, current time $t \geq s$, and no previous selection, one such policy stops if and only if $t \cdot q_t(s) \geq F_{t+1,s}$. Its optimal success probability is $\frac{1}{n}\sum_{s=1}^{n} F_{s,s}$. The optimal policy and value can be computed in time $O(n^2)$.

Proof sketch. Stopping before the signal or at a non-record cannot succeed. Conditional on $I = i$, the relative ranks are independent, and those before i are uniform on their respective ranges. At a feasible record state (t, s), Bayes' rule gives probabilities proportional to $tq_t(s)$ for $I = t$ and $q_i(s)$ for $I = i > t$: a record at t is certain in the first case and has probability $1/t$ in the second. The ranks before t have the same distribution in these cases, and future ranks are independent of the past conditional on I. Thus, at a record, only (t, s) matters. Backward induction yields an optimal deterministic policy that can be specified by a stopping set: accept the first record whose time belongs to that set.

For $t \geq s$ and a stopping set $A \subseteq \{t, \dots, n\}$, no record in A earlier than $i \in A$ is necessary for successfully stopping at i. Independence of the ranks conditional on $I = i$ allows us to write

$$F_{t,s} = \max_{A \subseteq \{t,\dots,n\}} \sum_{i \in A} q_i(s) \prod_{\substack{r \in A \\ r < i}} \left(1 - \frac{1}{r}\right).$$

For a given A, the sum divided by n is the probability of succeeding when observing $S = s$, rejecting before t, and then using A. Omitting t reduces to the best value $F_{t+1,s}$. Including t contributes $q_t(s)$ when $I = t$ and multiplies every later contribution by $1 - 1/t$, since a record at t (happening w.p. $1/t$) would cause the policy to stop too early. Comparing these choices leads to Equation (3).

At an observed record (t, s), the same calculation leads to conditional success probabilities $\frac{tq_t(s)}{\Psi_{t,s}}$ for accepting and $\frac{F_{t+1,s}}{\Psi_{t,s}}$ for rejecting and choosing optimally afterward. The denominator is the same: rejecting does not rule out $I = t$, but makes success impossible in that case, implying the stopping rule. Optimizing separately for each s and summing $\frac{F_{s,s}}{n}$ leads to the overall value. □

An optimal stopping set need not contain all times after a threshold: in Example 1 we give a four-item model with optimal stopping set $\{2, 4\}$ after observing $S = 2$. Stopping is the optimal choice at every feasible post-signal record state if and only if $q_t(s) \geq \sum_{u=t+1}^{n} \frac{q_u(s)}{u-1}$ for every feasible pair (t, s). This condition implies that $\mathcal{A}(S)$ is optimal (Theorem 10). The appendix also gives an exact criterion for optimality of $\mathcal{A}(S)$ under separable distributions (Theorem 11) and expresses its success probability for arbitrary distributions using only $\mathbb{E}[S \mid I = i]$ (Corollary 12).

2.3 Optimal policies and success probabilities for the α-power family

We now focus on the α-power family with $q_i(s) = g_s i^{-\alpha}$, where $g_s := s^\alpha - (s-1)^\alpha > 0$. Dividing Equation (3) by g_s yields $F_{t,s} = g_s f_t$, for $f_{n+1} = 0$ and $f_t = \max\{f_{t+1}, t^{-\alpha} + \frac{t-1}{t} f_{t+1}\}$. Thus, after the signal has arrived, its exact time no longer affects the decision at a record: Stopping when

$$t^{1-\alpha} \geq f_{t+1} \tag{4}$$

is an optimal rule. To find the stopping set, first compare accepting a record at t with rejecting it and accepting the first later record. By the stopping-set formula, accepting now is at least as good as this alternative exactly when $G_t \geq 0$, where $G_t := t^{1-\alpha} - t\sum_{u=t+1}^{n} \frac{1}{(u-1)u^\alpha}$. For $\alpha \geq 1$, summing $\frac{1}{(u-1)u^\alpha} \leq (u-1)^{-\alpha} - u^{-\alpha}$ over u implies $G_t \geq tn^{-\alpha} > 0$. Thus, by backward induction, accepting every post-signal record is optimal, yielding $\mathcal{A}(S)$. For $0 < \alpha < 1$ and $t \in [n]$,

$$\frac{G_{t+1}}{t+1} - \frac{G_t}{t} = \frac{(t+1)^{1-\alpha} - t^{1-\alpha}}{t} > 0\,.$$

Thus, G_t changes sign at most once. Let $k_n := \min\{t \in [n] : G_t \geq 0\}$. Backward induction verifies that accepting records from k_n onward is optimal after the signal. If $k_n > 1$, then $G_{k_n-1} = (k_n - 1)^{1-\alpha} - f_{k_n} < 0$, so rejecting is optimal at $k_n - 1$. Continuing backwards, rejecting keeps $f_t = f_{t+1}$, while $t^{1-\alpha}$ decreases, so rejecting remains optimal. This leads to $\mathcal{A}(\max\{S, k_n\})$.

For fixed $c \in (0,1]$ and $t = \lfloor cn \rfloor$, an integral approximation yields $n^{\alpha-1}G_t \to \frac{c-(1-\alpha)c^{1-\alpha}}{\alpha}$. Its root at $c = (1-\alpha)^{1/\alpha}$ determines the limiting threshold. Applying the same record calculation and integral approximation to the resulting policies yields the success probabilities summarized below.

Theorem 2. *For $\alpha \geq 1$, the policy $\mathcal{A}(S)$ is optimal. For $0 < \alpha < 1$, there is a threshold $k_n \in [n]$ such that the policy $\mathcal{A}(\max\{S, k_n\})$ is optimal and $\frac{k_n}{n} \xrightarrow{n\to\infty} (1-\alpha)^{1/\alpha}$. Let $\mathrm{OPT}_n(\alpha)$ denote the optimal success probability on instances of length n. Then*

$$\mathrm{OPT}_n(\alpha) \xrightarrow{n\to\infty} \mathrm{OPT}(\alpha) := \begin{cases} \dfrac{\alpha + (1-\alpha)^{1+1/\alpha}}{\alpha+1}, & 0 < \alpha < 1, \\ \dfrac{\alpha}{\alpha+1}, & \alpha \geq 1\,. \end{cases}$$

3 Random-order model: robustness and learning

The previous section determines an optimal policy for each α. We now study the effect of choosing a threshold using an estimate $\hat{\alpha}$ instead of the true α. As in the last section, the true signal remains in the α-power family, is reliable, and is independent of the relative order conditional on I.

A useful baseline requiring no parameter estimate is $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$, where $k_n^{\mathrm{cl}} \approx n/e$ is the classic optimal threshold. Whenever $k_n^{\mathrm{cl}} \leq S \leq I$, this policy succeeds on every realization where $\mathcal{A}(k_n^{\mathrm{cl}})$ succeeds: increasing its threshold from k_n^{cl} to S cannot introduce an earlier record. Thus, it preserves the finite-n success probability for any reliable signal, even with rank dependence. Tuning α aims to improve upon this baseline; we compare the two resulting policies in Table 2.

For a parameter estimate $\hat{\alpha}$, we first evaluate the policy $\mathcal{A}(\max\{S, \widetilde{k}_n(\beta)\})$ with threshold $\widetilde{k}_n(\beta) := \lceil \beta n \rceil$ for a fixed threshold fraction $\beta \in [0,1]$; $\beta = 0$ yields $\mathcal{A}(S)$. Similarly to the proof of Theorem 2, we obtain

$$\mathbb{P}\Big[\mathcal{A}(\max\{S, \widetilde{k}_n(\beta)\}) \text{ succeeds}\Big] \xrightarrow{n\to\infty} \frac{\alpha}{\alpha+1} + \frac{1-\alpha}{\alpha}\beta - \frac{\beta^{\alpha+1}}{\alpha(\alpha+1)} =: f(\alpha,\beta)\,. \tag{5}$$

Maximizing over β gives $\beta^\star(\alpha)$, where $\beta^\star(a) = (1-a)^{1/a}$ for $0 < a < 1$ and $\beta^\star(a) = 0$ for $a \geq 1$. The maximum is $\mathrm{OPT}(\alpha)$, so the rounded threshold is asymptotically optimal, though not necessarily the exact finite-n optimizer. When only an estimate $\hat{\alpha}$ is available, we use $\beta^\star(\hat{\alpha})$ instead.

Theorem 3. *Let $\alpha, \hat{\alpha} > 0$, where the true signal is α-power and the threshold is chosen using $\hat{\alpha}$. Set $\widehat{k}_n := \widetilde{k}_n(\beta^\star(\hat{\alpha})) = \lceil \beta^\star(\hat{\alpha}) n \rceil$. Then the policy $\mathcal{A}(\max\{S, \widehat{k}_n\})$ satisfies $\lim_{n\to\infty} \mathbb{P}[\textit{success}] = g(\alpha,\hat{\alpha}) := f(\alpha, \beta^\star(\hat{\alpha}))$. Explicitly,*

$$g(\alpha,\hat{\alpha}) = \begin{cases} \dfrac{\alpha}{\alpha+1} + \dfrac{1-\alpha}{\alpha}(1-\hat{\alpha})^{1/\hat{\alpha}} - \dfrac{(1-\hat{\alpha})^{(\alpha+1)/\hat{\alpha}}}{\alpha(\alpha+1)}, & 0 < \hat{\alpha} < 1, \\ \dfrac{\alpha}{\alpha+1}, & \hat{\alpha} \geq 1. \end{cases}$$

In particular, if $\hat{\alpha} \leq \alpha$, then $g(\alpha,\hat{\alpha}) \geq \mathrm{OPT}(\hat{\alpha}) \geq 1/e$. Moreover, for $0 < \alpha, \hat{\alpha} < 1$, let $D(\alpha,\hat{\alpha}) := \mathrm{OPT}(\alpha) - g(\alpha,\hat{\alpha})$. As $\delta \to 0$ subject to $0 < \alpha + \delta < 1$, we have $D(\alpha, \alpha+\delta) = C_\alpha \delta^2 + O_\alpha(|\delta|^3)$ for some constant $C_\alpha > 0$.

The conservative guarantee for $\hat{\alpha} \leq \alpha$ follows directly because larger α favors later signals. For a fixed threshold, delaying S cannot transform a success into a failure because $S \leq I$. Hence, $f(\alpha,\beta)$ is non-decreasing in α. Additionally respecting the classic threshold near n/e likewise cannot hurt, giving $\mathrm{OPT}(\hat{\alpha}) \geq 1/e$. For fixed $0 < \alpha < 1$, $f(\alpha,\cdot)$ has zero first derivative and negative second derivative at $\beta^\star(\alpha)$. Since $\beta^\star(\hat{\alpha})$ varies smoothly with $\hat{\alpha}$ and has nonzero derivative at α, Taylor expansion gives quadratic rather than linear loss. However, severe overestimation can lose the classic guarantee. If $\alpha, \hat{\alpha} \geq 1$, the policy is $\mathcal{A}(S)$ and is optimal even for finite n. But if $\hat{\alpha} \geq 1$ and $\alpha < \frac{1}{e-1}$, its limiting success probability is $\frac{\alpha}{\alpha+1} < \frac{1}{e}$ and tends to zero as $\alpha \to 0$. See the left panel of Figure 2; the full proof is in Appendix C.

Learning from historical instances. Under the same signal model, Theorem 16 constructs a lower confidence bound for α from independent historical observations of (I, S). With probability at least

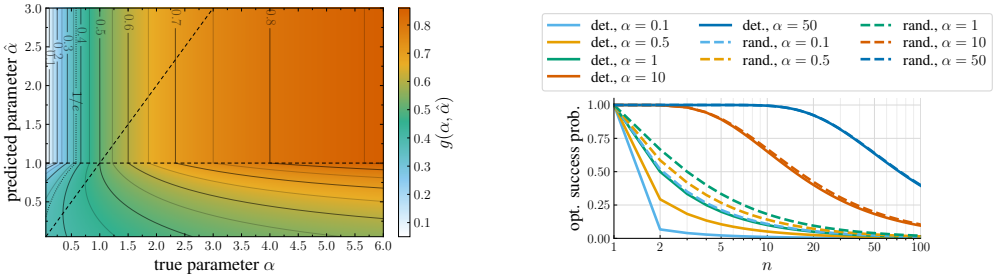

Figure 2: Left: Heatmap of the asymptotic guarantee $g(\alpha, \hat{\alpha})$ from Theorem 3. Right: Optimal worst-case success probabilities in adversarial order. Solid lines show the deterministic optimum $1 - (1 - 1/n)^{\alpha}$; dashed lines show the randomized optimum $\frac{n^{\alpha}}{\sum_{j=1}^{n} j^{\alpha}}$.

$1 - \delta$ over m informative instances ($I \geq 2$), its policy preserves the finite-n classic benchmark and has loss $O_{\alpha}(\sqrt{\log(2/\delta)/m})$ relative to the true optimum, simultaneously for every n. The middle panel of Figure 3 evaluates this policy as the training sample grows.

4 Adversarial Order

We consider an oblivious adversary that fixes an instance $v = (v_1, \ldots, v_n) \in \mathbb{N}_0^n$ with a unique maximum at $i^{\star}$ before the signal is drawn and the policy uses randomization. At time t, the policy observes v_t and any signal arrival before deciding whether to stop. Conditional on $i^{\star} = i$, the signal follows the α-power distribution on $[i]$. Let $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha)$ and $\mathrm{OPT}_n^{\mathrm{det}}(\alpha)$ denote the optimal randomized and deterministic worst-case success probabilities, respectively.

Randomized policies. Without a signal, the optimal randomized worst-case guarantee is $1/n$, attained by stopping at a uniformly random time. For $n \geq 2$, an additional α-power signal strictly improves this benchmark; we characterize the exact optimum below.

Theorem 4. *Let $\alpha > 0$. For adversarial arrival orders with an α-power signal, the optimal worst-case success probability over randomized policies is $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha) = \frac{n^{\alpha}}{\sum_{j=1}^{n} j^{\alpha}}$. The policy $\mathcal{A}(\max\{R, S\})$ is optimal, where $R \in [n]$ is independent of S and, for $r \in [n]$,*

$$\mathbb{P}[R \leq r] = \frac{n^{\alpha}}{\sum_{j=1}^{n} j^{\alpha}} \cdot \frac{\sum_{j=1}^{r} j^{\alpha}}{r^{\alpha}}.$$

The random threshold equalizes success on prefixes of $1, 2, \ldots, n$ padded with zeros. Their common prefixes and the one-choice constraint give a matching upper bound (Section D.1).

Corollary 5. *Let $(\alpha_n)_{n \geq 1}$ be a positive sequence. If $\alpha_n = o(n)$, then $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) = o(1)$. If $\alpha_n/n \to c \in (0, \infty)$, then $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) \to 1 - e^{-c}$.*

Deterministic policies. Without a signal, no deterministic policy has a positive worst-case success probability for $n \geq 2$. With an α-power signal, deterministic policies have a positive finite-n guarantee for every $\alpha > 0$:

Theorem 6. *Let $\alpha > 0$. For adversarial orders with an α-power signal, the policy $\mathcal{A}(S)$ is optimal among deterministic policies, and its worst-case success probability is $\mathrm{OPT}_n^{\mathrm{det}}(\alpha) = 1 - \left(1 - \frac{1}{n}\right)^{\alpha}$.*

Policy $\mathcal{A}(S)$ succeeds whenever $S = i^{\star}$. For positive worst-case success, the n signal times must cover all n possible maximum positions, forcing a deterministic policy to stop at the signal on these instances (Section D.2). The deterministic optimum has the same two limiting regimes as the randomized optimum.

Corollary 7. *Let $(\alpha_n)_{n \geq 1}$ be a positive sequence. If $\alpha_n = o(n)$, then $\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) = o(1)$. If $\alpha_n/n \to c \in (0, \infty)$, then $\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) \to 1 - e^{-c}$.*

5 MULTIPLE UNIFORM SIGNALS VERSUS ONE α-POWER SIGNAL

We next ask whether a policy can exploit the full history underlying an integer-power signal. Let $S_1, \dots, S_m$ be sampled independently and uniformly from $[i^\star]$, and let $L := \max_{j\in[m]} S_j$. Then, for $i \in [n]$ and $\ell \in [i]$, $\mathbb{P}[L \le \ell \mid i^\star = i] = \left(\frac{\ell}{i}\right)^m$, so revealing only L is exactly a single α-power signal with $\alpha = m$. The signals are unlabeled: in the full-history model, the policy observes their multiset of arrival times, equivalently the number c_t of signals arriving at each time $1 \le t \le i^\star$. All signals at time t are observed before that time's decision. A policy cannot succeed before all m signals have arrived: if a signal is still outstanding after time t, then $t < i^\star$.

Randomized policies. Let $\mathrm{OPT}_n^{\mathrm{rand,full}}(m)$ denote the optimal randomized worst-case success probability when all m signal times are observed.

Theorem 8. *For all integers $n, m \ge 1$, it holds that* $\mathrm{OPT}_n^{\mathrm{rand}}(m) = \mathrm{OPT}_n^{\mathrm{rand,full}}(m)$.

The reason is statistical sufficiency: for $h = (c_1, \dots, c_\ell)$ with $\sum_t c_t = m$, $c_\ell > 0$, and every $i \ge \ell$,

$$\mathbb{P}[h \mid L = \ell, i^\star = i] = \frac{m!/\prod_{t=1}^{\ell} c_t!}{\ell^m - (\ell-1)^m}. \tag{6}$$

This distribution does not depend on i. Once L is known, the remaining history supplies randomness without further information about the best item's position. A randomized policy can (randomly) generate that history after observing L and simulate any full-history policy, matching its success probability on every fixed arrival order. In general, a deterministic policy cannot create this auxiliary randomness, which explains the separation below.

Deterministic policies. For $n = 4$ and $m = 2$, the deterministic optimum with only the last signal is $\frac{7}{16}$ (Theorem 6). Once both signals arrive, sort their times as $a \le b$ and stop according to:

$$(1,1) \mapsto 1, \qquad (1,2) \mapsto 2, \qquad (1,3), (2,3), (3,3) \mapsto 3, \qquad \text{otherwise} \mapsto 4.$$

Counting both orders of the draws when $a < b$ gives success probabilities $1, \frac{1}{2}, \frac{5}{9}, \frac{1}{2}$ for maximum positions $1, 2, 3, 4$, respectively. These cover every instance because the rule ignores values, so its guarantee is $\min\{1, \frac{1}{2}, \frac{5}{9}, \frac{1}{2}\} = \frac{1}{2} > \frac{7}{16}$.

An integer linear program characterizes $\mathrm{OPT}_n^{\mathrm{det,full}}(m)$, the optimal deterministic worst-case success probability with all signals. A history $h = (c_1, \dots, c_n) \in \{0, \dots, m\}^n$ records c_t signals at each time t, with $\sum_{t=1}^n c_t = m$. Let $\ell(h) := \max\{t : c_t > 0\}$ be its last signal time and $\mathcal{H}_m$ the set of all such histories. When the maximum is at position i, history h has probability $\lambda_i(h) = \frac{m!}{c_1! \cdots c_{\ell(h)}!} \frac{1}{i^m}$ if $\ell(h) \le i$, and 0 otherwise. The factorial ratio counts assignments of the signals to the observed arrival times. Interpret $x_{h,t} = 1$ as mapping history h to stopping time $t \ge \ell(h)$. The first constraint chooses one time per history. The second constraint guarantees success probability at least z for each possible maximum position.

$$\begin{aligned} \max \quad z \quad \text{s.t.} \quad & \sum_{t=\ell(h)}^{n} x_{h,t} = 1 && \forall h \in \mathcal{H}_m \\ & \sum_{h \in \mathcal{H}_m : \ell(h) \le i} \lambda_i(h)\, x_{h,i} \ge z && \forall i \in [n] \\ & x_{h,t} \in \{0,1\} && \forall h \in \mathcal{H}_m,\ \forall t \in \{\ell(h), \dots, n\}. \end{aligned} \tag{ILP}$$

Theorem 9. *For all $m, n \in \mathbb{N}$, the optimal objective value of* (ILP) *is* $\mathrm{OPT}_n^{\mathrm{det,full}}(m)$.

Every deterministic policy induces such an assignment on $(1, \dots, n)$ and every feasible assignment defines a policy that ignores values and guarantees z on every instance (Section E.2). For $m = 2$ and $n \ge 4$, Theorem 18 and Corollary 19 give $\frac{6(n-1)}{(n+1)(2n+1)} \le \mathrm{OPT}_n^{\mathrm{det,full}}(2) \le \frac{6n}{(n+1)(2n+1)}$. Consequently, $\mathrm{OPT}_n^{\mathrm{det,full}}(2) = \frac{3}{n} + O(n^{-2})$, whereas observing only the last signal gives $\mathrm{OPT}_n^{\mathrm{det}}(2) = \frac{2}{n} - \frac{1}{n^2}$. Thus, the full history yields an asymptotic factor-$\frac{3}{2}$ improvement for deterministic policies, while it gives no improvement to randomized policies (Theorem 8).

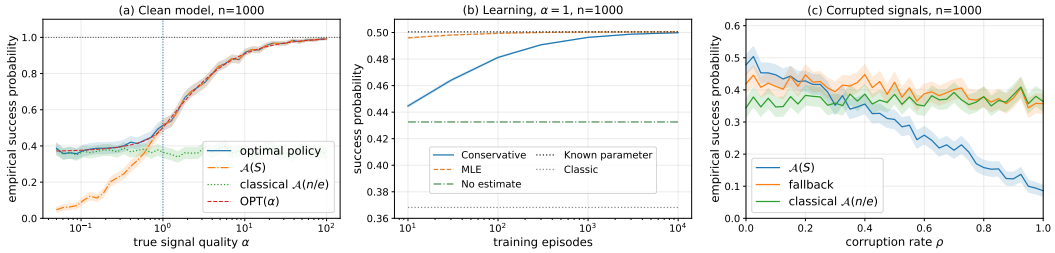

Figure 3: Random-order experiments at $n = 1000$. Left: clean-model empirical success. Middle: learning the signal parameter at $\alpha = 1$, with $\delta = 0.05$. Right: an out-of-model stress test with corrupted signals. Shading gives approximate 95% confidence intervals for the means.

6 Numerical Experiments

We complement the theory with synthetic experiments on clean, learned, and misspecified signals, and an out-of-model corruption stress test. Appendix F gives further experiments, including held-out fitting under distribution-family misspecification and rank dependence.

Setup. For the clean and misspecified comparisons, we sample uniformly random permutations, let I be the position of the maximum, and draw S from the α-power distribution conditional on I. Empirical success probabilities average 1000 runs. All comparisons use the exact finite-n classic threshold k_n^{cl}, abbreviated by n/e in the legends. The learning panel averages exact test success over sampled training histories.

Clean finite-sample behavior. The left panel of Figure 3 compares the finite-n optimal policy from Theorem 2, $\mathcal{A}(S)$, and the classic policy, with $\mathrm{OPT}(\alpha)$ as a reference. The predicted gain is already visible at $n = 1000$. For $\alpha \geq 1$, accepting the first record at or after the signal is optimal; for $\alpha < 1$, an additional deterministic threshold can improve success.

Learning the signal parameter. The middle panel compares the conservative estimator from Theorem 16, with $\delta = 0.05$, and MLE at $\alpha = 1$. Both use exact finite-n thresholds, fitted on the same 1000 independent histories at each training size. Baselines are the known-parameter optimum, the classic policy, and $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$, which needs no estimate. The conservative policy approaches the optimum; MLE learns faster on average, but does not guarantee conservativeness. Section F.3 gives results for $\alpha \in \{0.3, 1, 2\}$ and empirical coverage.

Corrupted asynchronous signals. The right panel fixes $\alpha = 1$ and corrupts the signal with probability ρ: it is missed, replaced by a uniformly random false alarm in $[n]$, or delayed to a uniformly random time after I, each with equal probability. When $I = n$, the delayed branch is unobserved. We compare $\mathcal{A}(S)$ with the classic policy and an online fallback: accept the first record at or after the signal; if no signal has arrived by k_n^{cl}, use the classic policy. Success declines more slowly for the fallback, but early false alarms prevent a classic-benchmark guarantee: at $\rho = 1$, its limiting success is $1/e - 1/(12e^2) < 1/e$ (Section F.1).

7 Discussion and Limitations

Our guarantees rely on assumptions about signal generation that may not hold in applications. The optimality results assume a known signal distribution satisfying $S \leq I$ and, in random order, independence of the signal and relative order conditional on I. Even a perfect fit to the signal distribution can conceal rank dependence that sharply reduces success (Section F.2), so fitting alone cannot validate the model. The robustness analysis covers only misspecification of α within the α-power family, not missing, false, or late signals. Our synthetic experiments explore these violations as stress tests, rather than validate the assumptions on real data. The mixture safeguard in Section F.1 provides a baseline, but optimal policies under corruption remain open. Further questions include sharper learning guarantees, validation of the signal model, and extensions to richer selection constraints.

AI USE STATEMENT

In this work, we used generative AI tools to review and edit the exposition, including alternative presentations of proofs, to help formulate and check model stress tests, and to assist with code used for experiments and figures. We did not use generative AI tools to originate the research question or theoretical model. The authors reviewed all AI-assisted work, checked the mathematical arguments, and verified the generated code and outputs. We take responsibility for the final content of this work, including text, claims, code, and figures.

ETHICS STATEMENT

This theoretical work uses only synthetic experiments and does not involve human subjects, personal data, or deployed systems.

REPRODUCIBILITY STATEMENT

Complete proofs are provided in the appendices. Section 6 and Appendix F describe the experimental setup, and the supplementary material includes the code and data used to produce the figures.

REFERENCES

Giovanni Abramo, Ciriaco Andrea D'Angelo, and Giovanni Felici. Predicting publication long-term impact through a combination of early citations and journal impact factor. *J. Informetrics*, 13(1): 32–49, 2019. doi: 10.1016/j.joi.2018.11.003.

Jonathan Adams. Early citation counts correlate with accumulated impact. *Scientometrics*, 63(3): 567–581, 2005. doi: 10.1007/s11192-005-0228-9.

Arpit Agarwal and Eric Balkanski. Learning-augmented dynamic submodular maximization. In *NeurIPS*, 2024. doi: 10.52202/079017-0453.

Spyros Angelopoulos, Marcin Bienkowski, Christoph Dürr, and Bertrand Simon. Contract scheduling with distributional and multiple advice. In *IJCAI*, pp. 3652–3660. ijcai.org, 2024. URL `https://www.ijcai.org/proceedings/2024/404`.

Antonios Antoniadis, Christian Coester, Marek Eliás, Adam Polak, and Bertrand Simon. Online metric algorithms with untrusted predictions. *ACM Trans. Algorithms*, 19(2):19:1–19:34, 2023a. doi: 10.1145/3582689.

Antonios Antoniadis, Themis Gouleakis, Pieter Kleer, and Pavel Kolev. Secretary and online matching problems with machine learned advice. *Discret. Optim.*, 48(Part 2):100778, 2023b. doi: 10.1016/j.disopt.2023.100778.

Pablo Daniel Azar, Robert Kleinberg, and S. Matthew Weinberg. Prior independent mechanisms via prophet inequalities with limited information. *Games Econ. Behav.*, 118:511–532, 2019. doi: 10.1016/j.geb.2018.05.006.

Yossi Azar, Stefano Leonardi, and Noam Touitou. Flow time scheduling with uncertain processing time. In *STOC*, pp. 1070–1080. ACM, 2021. doi: 10.1145/3406325.3451023.

Eric Balkanski, Will Ma, and Andreas Maggiori. Fair secretaries with unfair predictions. In *NeurIPS*, 2024. doi: 10.52202/079017-0122.

Nikhil Bansal, Christian Coester, Ravi Kumar, Manish Purohit, and Erik Vee. Learning-augmented weighted paging. In *SODA*, pp. 67–89. SIAM, 2022. doi: 10.1137/1.9781611977073.4.

M. J. Beckmann. Dynamic programming and the secretary problem. *Comput. Math. Appl.*, 19(11): 25–28, 1990. doi: 10.1016/0898-1221(90)90145-A.

Ziyad Benomar and Vianney Perchet. Advice querying under budget constraint for online algorithms. In *NeurIPS*, 2023.

Ziyad Benomar, Romain Cosson, Alexander Lindermayr, and Jens Schlöter. Non-clairvoyant scheduling with progress bars. In *NeurIPS*, 2025. URL https://openreview.net/forum?id=gYbreatcV1.

Hedyeh Beyhaghi and Linda Cai. Recent developments in pandora's box problem: Variants and applications. *SIGecom Exch.*, 21(1):20–34, 2023. doi: 10.1145/3699814.3699817.

Alexander Braun and Sherry Sarkar. The secretary problem with predicted additive gap. In *NeurIPS*, 2024. doi: 10.52202/079017-0521.

Davin Choo, Themistoklis Gouleakis, Chun Kai Ling, and Arnab Bhattacharyya. Online bipartite matching with imperfect advice. In *ICML*, 2024. URL https://proceedings.mlr.press/v235/choo24a.html.

José Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Recent developments in prophet inequalities. *SIGecom Exch.*, 17(1):61–70, 2018. doi: 10.1145/3331033.3331039.

José Correa, Andrés Cristi, Boris Epstein, and José A. Soto. The two-sided game of googol. *J. Mach. Learn. Res.*, 23:113:1–113:37, 2022a. URL https://jmlr.org/papers/v23/21-0630.html.

José Correa, Paul Dütting, Felix A. Fischer, and Kevin Schewior. Prophet inequalities for independent and identically distributed random variables from an unknown distribution. *Math. Oper. Res.*, 47(2):1287–1309, 2022b. doi: 10.1287/moor.2021.1167.

José Correa, Andrés Cristi, Laurent Feuilloley, Tim Oosterwijk, and Alexandros Tsigonias-Dimitriadis. The secretary problem with independent sampling. *Manag. Sci.*, 71(4):2778–2801, 2025. doi: 10.1287/mnsc.2021.01580.

Qiming Cui and Michael Dinitz. Ski rental with distributional predictions of unknown quality. *CoRR*, abs/2602.21104, 2026. doi: 10.48550/ARXIV.2602.21104.

Ilias Diakonikolas, Vasilis Kontonis, Christos Tzamos, Ali Vakilian, and Nikos Zarifis. Learning online algorithms with distributional advice. In *ICML*, volume 139, pp. 2687–2696. PMLR, 2021. URL http://proceedings.mlr.press/v139/diakonikolas21a.html.

Michael Dinitz, Sungjin Im, Thomas Lavastida, Benjamin Moseley, Aidin Niaparast, and Sergei Vassilvitskii. Binary search with distributional predictions. In *NeurIPS*, 2024. doi: 10.52202/079017-2871.

Paul Dütting, Silvio Lattanzi, Renato Paes Leme, and Sergei Vassilvitskii. Secretaries with advice. *Math. Oper. Res.*, 49(2):856–879, 2024. doi: 10.1287/moor.2023.1384.

Evgenii Borisovich Dynkin. The optimum choice of the instant for stopping a markov process. *Soviet Mathematics*, 4:627–629, 1963.

Marisa Dziallas and Knut Blind. Innovation indicators throughout the innovation process: An extensive literature analysis. *Technovation*, 80-81:3–29, 2019. ISSN 0166-4972. doi: 10.1016/j.technovation.2018.05.005. URL https://www.sciencedirect.com/science/article/pii/S0166497217301402.

Marek Eliás, Haim Kaplan, Yishay Mansour, and Shay Moran. Learning-augmented algorithms with explicit predictors. In *NeurIPS*, 2024. doi: 10.52202/079017-3109.

Hossein Esfandiari, MohammadTaghi Hajiaghayi, Brendan Lucier, and Michael Mitzenmacher. Prophets, secretaries, and maximizing the probability of choosing the best. In *AISTATS*, pp. 3717–3727. PMLR, 2020. URL http://proceedings.mlr.press/v108/esfandiari20a.html.

Thomas S. Ferguson. Who solved the secretary problem. *Stat. Sci.*, 4:282–289, 1989. doi: 10.1214/ss/1177012493.

Kaito Fujii and Yuichi Yoshida. The secretary problem with predictions. *Math. Oper. Res.*, 49(2): 1241–1262, 2024. doi: 10.1287/moor.2022.0031.

John P. Gilbert and Frederick Mosteller. Recognizing the maximum of a sequence. *J. Amer. Statistical Assoc.*, 61:35–73, 1966. doi: 10.1007/978-0-387-44956-2_22.

Anupam Gupta, Haim Kaplan, Alexander Lindermayr, Jens Schlöter, and Sorrachai Yingchareonthawornchai. A little clairvoyance is all you need. In *FOCS*, pp. 86–118. IEEE, 2025. doi: 10.1109/focs63196.2025.00010.

Theodore P. Hill and Robert P. Kertz. A survey of prophet inequalities in optimal stopping theory. *Contemporary mathematics*, 125:191–207, 1992. URL https://api.semanticscholar.org/CorpusID:2807359.

Ying Hu, Changjun Hu, Shushen Fu, Mingzhe Fang, and Wenwen Xu. Predicting key events in the popularity evolution of online information. *PLoS ONE*, 12, 2017. doi: 10.1371/journal.pone.0168749.

Zhihao Jiang, Pinyan Lu, Zhihao Gavin Tang, and Yuhao Zhang. Online selection problems against constrained adversary. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5002–5012. PMLR, 2021.

Billy Jin and Will Ma. Online bipartite matching with advice: Tight robustness-consistency tradeoffs for the two-stage model. In *NeurIPS*, 2022. doi: 10.52202/068431-1058.

Pooria Joulani, András György, and Csaba Szepesvári. Online learning under delayed feedback. In *ICML (3)*, pp. 1453–1461, 2013.

Yiming Kang. Secretary problem with predictions and ordering. In *AISTATS*, volume 300. PMLR, 2026. URL https://openreview.net/forum?id=jI1xC44cTn.

Haim Kaplan, David Naori, and Danny Raz. Competitive analysis with a sample and the secretary problem. *SIAM J. Comput.*, 54(6):1489–1513, 2025. doi: 10.1137/24m1650223.

Helia Karisani, Mohammadreza Daneshvaramoli, Hedyeh Beyhaghi, Mohammad Hajiesmaili, and Cameron Musco. The secretary problem with predictions and a chosen order. In *ITCS*, pp. 86:1–86:24. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2026. doi: 10.4230/LIPIcs.ITCS.2026.86.

Misha Khodak, Maria-Florina Balcan, Ameet Talwalkar, and Sergei Vassilvitskii. Learning predictions for algorithms with predictions. In *NeurIPS*, 2022. doi: 10.52202/068431-0256.

Silvio Lattanzi, Thomas Lavastida, Benjamin Moseley, and Sergei Vassilvitskii. Online scheduling via learned weights. In *SODA*, pp. 1859–1877. SIAM, 2020. doi: 10.1137/1.9781611975994.114.

Alexander Lindermayr and Nicole Megow. Permutation predictions for non-clairvoyant scheduling. *ACM Trans. Parallel Comput.*, 12(2):4:1–4:26, 2025. doi: 10.1145/3711872.

Alexander Lindermayr and Nicole Megow. Repository of papers on algorithms with predictions, 2026. http://algorithms-with-predictions.github.io/.

Dennis V. Lindley. Dynamic programming and decision theory. *J. Roy. Stat. Soc. C-app.*, 10:39–51, 1961. doi: 10.2307/2985407.

Xujun Liu, Olgica Milenkovic, and George V. Moustakides. Query-based selection of optimal candidates under the mallows model. *Theor. Comput. Sci.*, 979:114206, 2023.

Brendan Lucier. An economic view of prophet inequalities. *SIGecom Exch.*, 16(1):24–47, 2017. doi: 10.1145/3144722.3144725.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *J. ACM*, 68(4):24:1–24:25, 2021. doi: 10.1145/3447579.

Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *Commun. ACM*, 65(7):33–35, 2022. doi: 10.1145/3528087.

Benjamin Moseley, Heather Newman, Kirk Pruhs, and Rudy Zhou. Robust gittins for stochastic scheduling. In *SIGMETRICS (Abstracts)*, pp. 166–168. ACM, 2025. doi: 10.1145/3744970.3727315.

George V. Moustakides, Xujun Liu, and Olgica Milenkovic. Optimal stopping methodology for the secretary problem with random queries. *J. Appl. Probab.*, 61(2):578–602, 2024.

Hasti Nourmohammadi, Ying Cao, Bo Sun, and Xiaoqi Tan. Ordinal secretaries with advice. In *AAAI*, volume 40, pp. 37108–37116, 2026. doi: 10.1609/aaai.v40i43.41040.

Pranav Nuti. The secretary problem with distributions. In *IPCO*, pp. 429–439. Springer, 2022. doi: 10.1007/978-3-031-06901-7_32.

Pranav Nuti and Jan Vondrák. Secretary problems: The power of a single sample. In *SODA*, pp. 2015–2029. SIAM, 2023.

Yu-Yang Qian, Zhen-Yu Zhang, Peng Zhao, and Zhi-Hua Zhou. Learning with asynchronous labels. *ACM Trans. Knowl. Discov. Data*, 18(8):186:1–186:27, 2024. doi: 10.1145/3662186.

Vinod Raman and Ambuj Tewari. Online classification with predictions. In *NeurIPS*, 2024. doi: 10.52202/079017-1777.

Aviad Rubinstein, Jack Z. Wang, and S. Matthew Weinberg. Optimal single-choice prophet inequalities from samples. In *ITCS*, pp. 60:1–60:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi: 10.4230/LIPICS.ITCS.2020.60.

Fabian Spaeh and Alina Ene. Online ad allocation with predictions. In *NeurIPS*, 2023. doi: 10.52202/075280-0756.

Gábor Szabó and Bernardo A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, 2010. doi: 10.1145/1787234.1787254.

Alexandru-Florin Tatar, Marcelo Dias de Amorim, Serge Fdida, and Panayotis Antoniadis. A survey on predicting the popularity of web content. *J. Internet Serv. Appl.*, 5(1):8:1–8:20, 2014. doi: 10.1186/s13174-014-0008-y.

Alexander Wei. Better and simpler learning-augmented online caching. In *APPROX-RANDOM*, volume 176 of *LIPIcs*, pp. 60:1–60:17, 2020. doi: 10.4230/LIPICS.APPROX/RANDOM.2020.60.

Martin Weitzman. Optimal search for the best alternative. *Econometrica*, 47:641–654, 1978. URL `https://api.semanticscholar.org/CorpusID:32530881`.

Jia-Qi Yang and De-Chuan Zhan. Generalized delayed feedback model with post-click information in recommender systems. In *NeurIPS*, 2022.

A Further related work

Dynamic programming is a classic tool for secretary problems (Lindley, 1961; Beckmann, 1990); our Bellman recursion applies it to the posterior induced by a stochastic lower bound on the optimum's position. Table 1 summarizes representative nearby models.

Relation to the general advice framework. Dütting et al. (2024) allow a known joint distribution of ranks and per-arrival advice. Encode our observation at time t as $X_t = 0$ if $t < S$ and $X_t = S$ otherwise. This satisfies their non-filtering and history-irrelevance conditions. Indeed, X_t determines all earlier advice. Conditional on a record at t, the earlier relative ranks are independent uniforms, both when $I = t$ and when $I > t$, and conditioning further on the signal does not change them. Thus, future record/advice observations do not change the earlier history conditional on its current record/advice state (non-filtering), and the past history gives no further information about $I = t$ (history irrelevance). If $X_t = 0$, then $I > t$ is already certain. Consequently, their exact LP and memoryless-policy characterization apply. Theorem 1 gives an explicit $O(n^2)$ implementation for this signal model; our subsequent results derive its threshold conditions, closed forms, and adversarial guarantees.

Table 1: A taxonomy of representative secretary models with auxiliary information. "Before arrivals" means before the online sequence starts. The general framework of Dütting et al. (2024) also contains our random-order model.

Model	When available	Advice and uncertainty
Sample-based secretary Nuti & Vondrák (2023); Kaplan et al. (2025); Correa et al. (2025)	Before arrivals	Values of a random, nonselectable sample; uncertainty comes from sampling.
Value predictions Jiang et al. (2021); Antoniadis et al. (2023b); Fujii & Yoshida (2024); Balkanski et al. (2024); Karisani et al. (2026); Kang (2026)	Before arrivals	Predicted maximum or candidate values; predictions may be arbitrary, with performance parameterized by model-specific errors.
Aggregate or ordinal advice Braun & Sarkar (2024); Nourmohammadi et al. (2026)	Before arrivals	Predicted value gap or position of the best candidate; evaluated by absolute error, consistency, or average robustness.
General stochastic advice Dütting et al. (2024)	With each candidate	Known joint rank/signal distribution; includes noisy binary recommendations and our signal encoding.
Queried expert Liu et al. (2023); Benomar & Perchet (2023); Moustakides et al. (2024)	During arrivals	A limited number of adaptively chosen queries to a perfect or noisy expert.
This work	During arrivals; equivalent upfront	A stochastic lower bound on the best item's position, with a known conditional signal distribution.

Relation to positional predictions. Nourmohammadi et al. (2026) study an upfront, potentially erroneous prediction of the best candidate's position and consistency-robustness tradeoffs. Our signal is also positional advice and can equivalently be revealed upfront. Its additional assumptions are a reliable lower bound and a known conditional distribution. These permit optimization of the exact expected success probability under the specified joint model and lead to the explicit policy and value formulas studied here.

A natural extension of the classic secretary problem is the *full information* setting, where item values are drawn i.i.d. from a known distribution (Gilbert & Mosteller, 1966; Esfandiari et al., 2020; Nuti, 2022). More broadly, optimal stopping problems such as prophet inequalities (Correa et al., 2018; Lucier, 2017; Hill & Kertz, 1992), Pandora's box problems (Beyhaghi & Cai, 2023; Weitzman, 1978), and related variants typically assume full distributional knowledge. Recent work relaxes this assumption by providing samples from the underlying distributions, both in prophet (Azar et al., 2019; Correa et al., 2022a;b; Rubinstein et al., 2020) and secretary (Nuti & Vondrák, 2023; Kaplan et al., 2025; Dütting et al., 2024; Correa et al., 2025) settings, or by assuming only partial or approximate distributional advice (Diakonikolas et al., 2021; Cui & Dinitz, 2026; Dinitz et al., 2024; Angelopoulos et al., 2024; Moseley et al., 2025). These models provide information about item values or their distributions. Our signaling model carries only temporal information tied to the arrival time of the optimum and cannot be used to evaluate any candidate directly.

In the *learning-augmented algorithms* framework (Mitzenmacher & Vassilvitskii, 2022; Lindermayr & Megow, 2026), algorithms receive a potentially erroneous machine-learned prediction, and performance is typically expressed as a function of prediction error. This paradigm has been studied for caching (Lykouris & Vassilvitskii, 2021; Wei, 2020; Bansal et al., 2022; Antoniadis et al., 2023a; Eliás et al., 2024), scheduling (Lattanzi et al., 2020; Azar et al., 2021); (Lindermayr & Megow, 2025), matching and allocation (Jin & Ma, 2022; Spaeh & Ene, 2023; Choo et al., 2024), submodular maximization (Agarwal & Balkanski, 2024), and online learning (Khodak

et al., 2022; Raman & Tewari, 2024). In optimal stopping, predictions typically concern the rank or value of the current item and, if perfect, directly identify when to stop (Jiang et al., 2021; Antoniadis et al., 2023b; Fujii & Yoshida, 2024; Braun & Sarkar, 2024; Balkanski et al., 2024; Nourmohammadi et al., 2026; Karisani et al., 2026; Kang, 2026). A complementary line lets the algorithm decide when to query a perfect or noisy expert, subject to a query budget (Liu et al., 2023; Benomar & Perchet, 2023; Moustakides et al., 2024). The signal instead supplies a probabilistic but trustworthy temporal clue without being requested by the algorithm.

Closest in spirit is recent work on non-clairvoyant scheduling (Gupta et al., 2025; Benomar et al., 2025), where a delayed external signal provides partial information about each job's characteristics. Our signaling model shares the flavor of delayed and reliable side information, but is tied to the optimality of an item rather than to an intrinsic item feature such as processing time.

B Omitted proofs from Section 2

This section gives the formal proofs omitted from Section 2. We first prove the recursion for arbitrary signal distributions and its corollaries, and then specialize it to the α-power family.

B.1 Arbitrary distributions

We continue under the random-order signal model from Section 2: the items have distinct values, I is uniform on $[n]$, and the known signal distribution $q_i(s) = \mathbb{P}[S = s \mid I = i]$ is supported on $s \le i$. Conditional on I, the signal is independent of the relative order and is observed before the decision at time S. Recall the posterior normalizer $\Psi_{t,s} = tq_t(s) + \sum_{i=t+1}^{n} q_i(s)$, and that (t, s) is feasible if $\Psi_{t,s} > 0$.

Theorem 1. *For each $s \in [n]$, set $F_{n+1,s} := 0$ and, for $t = n, n-1, \dots, s$, define*

$$F_{t,s} := \max\left\{ F_{t+1,s},\, q_t(s) + \left(1 - \frac{1}{t}\right) F_{t+1,s} \right\}. \tag{3}$$

There is an optimal deterministic policy that never stops before the signal and only stops at records. At every feasible record state (t, s) with observed signal $S = s$, current time $t \ge s$, and no previous selection, one such policy stops if and only if $t \cdot q_t(s) \ge F_{t+1,s}$. Its optimal success probability is $\frac{1}{n}\sum_{s=1}^{n} F_{s,s}$. The optimal policy and value can be computed in time $O(n^2)$.

Proof. We prove the theorem in five steps.

Step 1: Structural reduction and posterior state. Before the signal arrives, the best item cannot have arrived because $S \le I$. A non-record cannot be the best item either, since a better item has already appeared. Hence, we may restrict attention to policies that stop only at records in $\{S, \dots, n\}$.

We denote by R_r the relative rank at time r, with $R_r = 1$ meaning that time r is a record time. To generate the ranking of the items from best to worst, insert the item arriving at time r into one of the r positions in the ranking of the first $r-1$ items, chosen uniformly and independently of earlier choices. Each final ranking arises from exactly one sequence of insertion positions and thus has probability $\prod_{r=1}^{n} \frac{1}{r} = \frac{1}{n!}$. This generates a uniformly random permutation. Since the insertion position at time r is exactly R_r, the variables $R_1, \dots, R_n$ are independent and each R_r is uniform on $[r]$.

The best item arrives at time i exactly when time i is the final record. Therefore, we have

$$\{I = i\} = \{R_i = 1\} \cap \bigcap_{r=i+1}^{n} \{R_r > 1\}.$$

Conditioning on $I = i$ therefore leaves all ranks before i unrestricted, fixes $R_i = 1$, and requires $R_r > 1$ at each later time r. Because these conditions concern individual ranks and the ranks are independent, they remain independent after conditioning. Thus, R_r is still uniform on $[r]$ for $r < i$, and is uniform on $\{2, \dots, r\}$ for $r > i$. In particular, once $I = i$ is fixed, observing past relative ranks gives no further information about future relative ranks.

Fix a feasible pair (t, s) and a particular observed rank sequence $\sigma_t = (R_1, \ldots, R_t)$ with $R_t = 1$. Let $\mathcal{H}_t := (\sigma_t, S = s)$ denote the resulting post-signal record history. If $i > t$, the first t ranks remain independent and uniform, so the probability of observing σ_t is $\prod_{r=1}^{t} \frac{1}{r} = \frac{1}{t!}$. If $i = t$, the record at t is certain, so only the first $t-1$ ranks contribute to this probability, giving $\frac{1}{t-1}!$. If $i < t$, the observed record at t is impossible because no record can occur after the best item has arrived. To obtain the joint probability below, we multiply each of these probabilities by $\mathbb{P}[I = i] = \frac{1}{n}$ and by the signal probability $q_i(s)$, using the independence of the signal and the relative ranks conditional on $I = i$:

$$\mathbb{P}[\mathcal{H}_t, I = i] = \begin{cases} \dfrac{q_t(s)}{n \cdot (t-1)!}, & i = t, \\ \dfrac{q_i(s)}{n \cdot t!}, & i > t, \\ 0, & i < t. \end{cases}$$

Summing over i gives the history probability

$$\mathbb{P}[\mathcal{H}_t] = \frac{1}{n \cdot t!} \left(t \cdot q_t(s) + \sum_{i=t+1}^{n} q_i(s) \right) = \frac{\Psi_{t,s}}{n \cdot t!}.$$

Thus, $\Psi_{t,s} > 0$ is exactly the condition that this record history has positive probability. Bayes' rule now gives

$$\mathbb{P}[I = t \mid \mathcal{H}_t] = \frac{t \cdot q_t(s)}{\Psi_{t,s}} \text{ and } \mathbb{P}[I = i \mid \mathcal{H}_t] = \frac{q_i(s)}{\Psi_{t,s}} \quad \text{for every } i > t. \tag{7}$$

We next justify why no other part of the relative-rank history is relevant. Let $\mathcal{B}$ be any event determined by the future ranks $R_{t+1}, \ldots, R_n$. For every $i \geq t$ with positive posterior probability, the independence of past and future ranks conditional on $I = i$ combined with the conditional independence of the signal from all ranks yields

$$\mathbb{P}[\mathcal{B} \mid I = i, \mathcal{H}_t] = \mathbb{P}[\mathcal{B} \mid I = i].$$

Hence,

$$\mathbb{P}[I = i, \mathcal{B} \mid \mathcal{H}_t] = \mathbb{P}[I = i \mid \mathcal{H}_t] \cdot \mathbb{P}[\mathcal{B} \mid I = i].$$

By Equation (7), the posterior weights depend only on (t, s), and the conditional future-rank distribution given $I = i$ does not depend on σ_t. Thus, the entire conditional joint distribution of I and the future ranks is the same for every post-signal record history with the same pair (t, s).

At such a history, the *remaining stopping problem* is to select at most one item from times $t, \ldots, n$, assuming that no item has previously been selected and that the current record is still available. Decisions must be based only on the observations available at that time, and the objective is the conditional probability of selecting the best item. Any admissible rule at one history can be used at another history with the same (t, s) by making the same decisions for each successive future-rank prefix, with the same independent internal randomization if any. This preserves both the online constraint and the conditional success probability since the joint distribution just established is identical. Consequently, the optimal value depends only on (t, s), and we may write

$$\Pi_{t,s} := \max_{\tau} \mathbb{P}[\tau = I \mid \mathcal{H}_t],$$

where the maximum is over online stopping rules with $\tau \in \{t, \ldots, n\} \cup \{\infty\}$, and $\tau = \infty$ means that no item is selected. Importantly, $\Pi_{t,s}$ is the value *before* deciding at time t: stopping now is allowed. We reserve $\Pi_{t,s}^{\text{cont}}$ for the value when the current record is rejected and only later items may be selected.

We now show by backward induction that optimal actions can be chosen simultaneously at all record histories using only (t, s). At $t = n$, stopping wins at every feasible record history. Suppose optimal decisions have been chosen at all later record times. If the current record is rejected, all following non-records until the next record must also be rejected. At the next record time $u > t$, the induction hypothesis supplies an optimal rule with value $\Pi_{u,s}$, independently of the in-between rank history; if no further record occurs, success is impossible. Since the joint distribution of I and the future

ranks depend only on (t, s), the stopping value as well as the value of rejecting the current record and acting optimally afterward depend only on (t, s). A randomized choice is a convex combination of these two values, so choosing a maximizing pure action, with ties resolved in favor of stopping, completes the induction. Thus, an optimal deterministic policy exists whose record-state decisions depend only on (t, s).

For each observed signal time s, let $A_s \subseteq \{s, \dots, n\}$ contain exactly those times t at which this policy stops upon encountering a record. Conversely, any family $(A_s)_{s\in[n]}$ defines a policy that accepts the first record whose time lies in the corresponding set A_s.

Step 2: Success probability of a fixed stopping set. Fix a signal time s and a stopping set $A_s \subseteq \{s, \dots, n\}$, which the policy uses when $S = s$. Conditional on $I = i$, time i contains the best item and is therefore a record. On the event $S = s$, the policy succeeds exactly when $i \in A_s$ and no earlier time $r \in A_s$ with $r < i$ is a record. Indeed, a record at any such earlier time would make the policy stop before i.

By Step 1, conditioned on $I = i$, the relative ranks before i are independent, with the relative rank R_r uniform on $[r]$. Thus, the probability that an earlier time $r < i$ is not a record is $\mathbb{P}[R_r > 1 \mid I = i] = 1 - \frac{1}{r}$. The signal is also independent of these ranks conditioned on $I = i$, so we can multiply the signal probability by the probabilities of having no record at the earlier times in A_s:

$$\begin{aligned}\mathbb{P}[\text{success}, S = s \mid I = i] &= \mathbb{1}_{\{i\in A_s\}}\mathbb{P}[S = s \mid I = i] \prod_{\substack{r\in A_s\\ s\le r<i}} \mathbb{P}[R_r > 1 \mid I = i] \\ &= q_i(s)\mathbb{1}_{\{i\in A_s\}} \prod_{\substack{r\in A_s\\ s\le r<i}} \left(1 - \frac{1}{r}\right).\end{aligned}$$

For the recursion in the next step, we use the same calculation for stopping sets restricted to times $t, \dots, n$, where $t \ge s$. For $A \subseteq \{t, \dots, n\}$, define

$$W_{t,s}(A) := \sum_{i\in A} q_i(s) \prod_{\substack{r\in A\\ t\le r<i}} \left(1 - \frac{1}{r}\right). \tag{8}$$

Since each value of I has probability $\frac{1}{n}$, averaging the success probabilities computed above gives

$$\mathbb{P}[\text{success}, S = s] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}[\text{success}, S = s \mid I = i] = \frac{1}{n} W_{s,s}(A_s).$$

Step 3: Dynamic-programming recursion. Fix a signal time s. We show that the values $F_{t,s}$ defined by Equation (3) satisfy, for every $t \in \{s, \dots, n\}$,

$$F_{t,s} = \max_{A\subseteq\{t,\dots,n\}} W_{t,s}(A). \tag{9}$$

We prove the claim by backward induction on t. For $t = n$, the only choices are $A = \emptyset$ and $A = \{n\}$, with values 0 and $q_n(s)$, respectively. Their maximum is $q_n(s)$. The recursion gives the same value $F_{n,s} = q_n(s)$ because $F_{n+1,s} = 0$.

Now let $s \le t < n$ and assume the claim holds at $t + 1$. We consider separately the stopping sets that omit t and those that include t.

If $t \notin A$, then $A \subseteq \{t + 1, \dots, n\}$ and $W_{t,s}(A) = W_{t+1,s}(A)$. Thus, the best value among these sets is $F_{t+1,s}$ by the induction hypothesis.

If $t \in A$, write $A = \{t\} \cup B$ with $B \subseteq \{t + 1, \dots, n\}$. When $I = t$, the policy succeeds whenever $S = s$ since no earlier time belongs to A. This gives the term $q_t(s)$. When $I = i \in B$, success requires that time t is not a record; otherwise, the policy would stop before i. Conditional on $I = i$, this has probability $1 - \frac{1}{t}$. Thus, compared with $W_{t+1,s}(B)$, each term for a later best item gains the factor $1 - \frac{1}{t}$, while all its other factors remain unchanged. Therefore,

$$W_{t,s}(\{t\} \cup B) = q_t(s) + \left(1 - \frac{1}{t}\right) W_{t+1,s}(B).$$

Maximizing over B and using the induction hypothesis, the best value among sets containing t is

$$q_t(s) + \left(1 - \frac{1}{t}\right) F_{t+1,s}.$$

Every stopping set falls into one of these two cases. Taking the larger of their best values gives exactly Equation (3), completing the induction and proving Equation (9).

Step 4: Optimal action at a feasible record state. Fix a feasible record history $\mathcal{H}_t$ with signal $S = s$ and no previous selection. We compare accepting the current record with rejecting it and choosing optimally among later items.

Stopping succeeds exactly when $I = t$. Hence, by Equation (7),

$$\Pi^{\text{stop}}_{t,s} = \frac{t \cdot q_t(s)}{\Psi_{t,s}}.$$

Suppose instead that the policy rejects the current record. By Step 1, optimal decisions at later records can be chosen to depend only on the time and s, so it suffices to consider stopping sets $A \subseteq \{t+1, \dots, n\}$.

Fix such a set A. If $I = t$, the best item has been rejected and success is impossible. For each $i > t$, the probability of $I = i$ given $\mathcal{H}_t$ is $\frac{q_i(s)}{\Psi_{t,s}}$. Conditional on $I = i$, the policy succeeds exactly when $i \in A$ and every earlier time in A is a non-record. By Step 1, these future ranks remain independent and uniform when we also condition on $\mathcal{H}_t$. Therefore,

$$\mathbb{P}[\text{success} \mid \mathcal{H}_t, \text{ continue and then use } A] = \sum_{i=t+1}^{n} \frac{q_i(s)}{\Psi_{t,s}} \mathbb{1}_{\{i \in A\}} \prod_{\substack{r \in A \\ t<r<i}} \left(1 - \frac{1}{r}\right) = \frac{W_{t+1,s}(A)}{\Psi_{t,s}}.$$

Rejecting the current record does not tell us that $I > t$: the item just rejected could still be the best one. We therefore keep the denominator $\Psi_{t,s}$ rather than condition on $I > t$. Maximizing over A and using Equation (9) gives

$$\Pi^{\text{cont}}_{t,s} = \frac{F_{t+1,s}}{\Psi_{t,s}}.$$

Consequently,

$$\Pi_{t,s} = \max\left\{\Pi^{\text{stop}}_{t,s}, \Pi^{\text{cont}}_{t,s}\right\} = \frac{\max\{t \cdot q_t(s),\, F_{t+1,s}\}}{\Psi_{t,s}}.$$

Since $\Psi_{t,s} > 0$, stopping is optimal exactly when $t \cdot q_t(s) \geq F_{t+1,s}$. This is the same condition under which including t in the stopping set is at least as good as omitting it in Step 3:

$$q_t(s) + \left(1 - \frac{1}{t}\right) F_{t+1,s} \geq F_{t+1,s} \quad \Longleftrightarrow \quad t \cdot q_t(s) \geq F_{t+1,s}.$$

Thus, for each s, define

$$A^*_s := \left\{t \in \{s, \dots, n\} : t \cdot q_t(s) \geq F_{t+1,s}\right\}.$$

Starting with the empty set after time n and working backwards, Step 3 shows that $A^*_s \cap \{t, \dots, n\}$ has value $F_{t,s}$ for every $t \in \{s, \dots, n\}$: at each time t, its definition chooses whichever of including or omitting t gives the larger value. Hence, the policy specified by the sets $(A^*_s)_{s \in [n]}$ is optimal and obeys the stated action rule at every feasible record state. If $t \cdot q_t(s) > F_{t+1,s}$, stopping is the only optimal action. With equality, either action is optimal, and our policy stops. States with $\Psi_{t,s} = 0$ cannot occur, so the actions specified there do not matter.

Step 5: Overall value and running time. Because S is observed, the policy can choose A_s separately for each signal time s. The total success probability is the sum of the probabilities of succeeding with each possible signal time. Each term $W_{s,s}(A_s)$ depends only on A_s, so we can maximize these terms separately. Using Step 2 and Equation (9), the success probability of an optimal policy is

$$\max_{(A_1, \dots, A_n)} \frac{1}{n} \sum_{s=1}^{n} W_{s,s}(A_s) = \frac{1}{n} \sum_{s=1}^{n} \max_{A_s \subseteq \{s, \dots, n\}} W_{s,s}(A_s)$$

$$= \frac{1}{n} \sum_{s=1}^{n} F_{s,s}.$$

This proves the value formula in Theorem 1. There is one recursively computed entry for every pair $1 \le s \le t \le n$, hence $\frac{n(n+1)}{2}$ such entries, in addition to the initialized boundary values $F_{n+1,s} = 0$. Each entry $F_{t,s}$ is computed from $F_{t+1,s}$ and $q_t(s)$ using a constant number of arithmetic operations and one comparison. The same comparison determines whether the policy stops or continues at a record at time t after observing $S = s$. Thus, if the signal probabilities are given in a table and each arithmetic operation takes constant time, the full policy and value can be computed in $O(n^2)$ time. □

The following example shows that the optimal policy may not be a single threshold.

Example 1. Let $n = 4$ and consider the signal model

$$q_1(1) = q_2(2) = q_3(1) = q_4(2) = 1,$$

with all other entries equal to zero. Thus, $S = 2$ occurs exactly when $I = 2$ or $I = 4$. Since I is uniform on $[4]$, we have $\mathbb{P}[S = 2] = \frac{1}{2}$. Starting from $F_{5,2} = 0$, the recursion gives

$$F_{4,2} = 1, \qquad F_{3,2} = 1, \qquad F_{2,2} = \frac{3}{2}.$$

The stopping rule compares $t \cdot q_t(2)$ with $F_{t+1,2}$. At times 2, 3, and 4, these comparisons are $2 > 1$, $0 < 1$, and $4 > 0$, respectively. Hence, it is strictly better to stop at records at times 2 and 4, but to reject a record at time 3. Thus, the optimal stopping set is $A_2^* = \{2, 4\}$.

For any stopping set A used after $S = 2$, Step 2 gives $\mathbb{P}[\text{success}, S = 2] = \frac{W_{2,2}(A)}{4}$. Dividing by $\mathbb{P}[S = 2] = \frac{1}{2}$ shows that its conditional success probability is $\frac{W_{2,2}(A)}{2}$. For A_2^*, this is $\frac{3/2}{2} = \frac{3}{4}$.

By contrast, a threshold policy accepts the first record at or after its threshold. Its nonempty stopping set after $S = 2$ must therefore be $\{4\}$, $\{3, 4\}$, or $\{2, 3, 4\}$. Their values of $W_{2,2}$ are 1, $\frac{2}{3}$, and

$$1 + \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right) = \frac{4}{3},$$

respectively. Hence, the best threshold uses $\{2, 3, 4\}$ and has conditional success probability $\frac{4/3}{2} = \frac{2}{3} < 3/4$. Thus, no threshold policy is optimal for this signal model.

Theorem 10. *Under the assumptions of Theorem 1, stopping is optimal at every feasible post-signal record state if and only if*

$$q_t(s) \ge \sum_{u=t+1}^{n} \frac{q_u(s)}{u-1}, \tag{10}$$

for every feasible pair (t, s). In particular, this condition implies that $\mathcal{A}(S)$ is optimal.

This statement requires stopping to be optimal at every feasible record state, including states that $\mathcal{A}(S)$ never reaches because it would have stopped earlier. The policy's decisions at those states do not affect its success probability. Thus, $\mathcal{A}(S)$ can be optimal even if stopping is not optimal at every feasible state.

Proof. Fix a signal time s. If $q_i(s) = 0$ for every $i \ge s$, then $S = s$ is impossible and there is nothing to prove for this signal time. Otherwise, let

$$m_s := \max\{i \ge s : q_i(s) > 0\}.$$

Thus, m_s is the latest possible arrival time of the best item when $S = s$. For $s \le t < m_s$, we have $\Psi_{t,s} \ge q_{m_s}(s) > 0$. At $t = m_s$, we have $\Psi_{m_s,s} = m_s q_{m_s}(s) > 0$, whereas for $t > m_s$, we have $\Psi_{t,s} = 0$. Hence, the feasible states with signal s are exactly (t, s) for $s \le t \le m_s$. A record at time m_s must be the best item, so stopping there succeeds with probability 1.

Sufficiency. Assume Equation (10) holds at every feasible state with signal s. We use backward induction, starting from $t = m_s$, where stopping is optimal as shown above. Now fix $s \le t < m_s$,

let t be a record time and assume stopping is optimal at every later feasible record state. If we reject the record at t, it is therefore optimal to accept the first later record, using the stopping set $\{t+1, \dots, m_s\}$. By Equation (9), we obtain

$$\begin{aligned} F_{t+1,s} &= W_{t+1,s}(\{t+1, \dots, m_s\}) \\ &= \sum_{u=t+1}^{m_s} q_u(s) \prod_{r=t+1}^{u-1} \frac{r-1}{r} \\ &= t \sum_{u=t+1}^{m_s} \frac{q_u(s)}{u-1} = t \sum_{u=t+1}^{n} \frac{q_u(s)}{u-1}, \end{aligned}$$

The product reduces to $\frac{t}{u-1}$ by cancelling consecutive factors, and the last equality uses $q_u(s) = 0$ for $u > m_s$. Multiplying the assumed inequality Equation (10) by t gives $t \cdot q_t(s) \geq F_{t+1,s}$. By Theorem 1, stopping is therefore optimal at (t, s). This completes the induction.

Necessity. Conversely, suppose stopping is optimal at every feasible state with signal s. Fix $t \in \{s, \dots, m_s\}$. After rejecting the record at t, it is then optimal to accept the first later record. The same calculation as above gives

$$F_{t+1,s} = t \sum_{u=t+1}^{n} \frac{q_u(s)}{u-1}.$$

Optimality of stopping at (t, s) requires $t \cdot q_t(s) \geq F_{t+1,s}$. Substituting the expression above and dividing by $t > 0$ gives Equation (10). Applying the argument to every signal time s proves the equivalence. Finally, stopping at every record at or after the signal is exactly $\mathcal{A}(S)$, so the condition also implies that this policy is optimal. □

Theorem 11. *Let $a_1, \dots, a_n > 0$, set $A_i := \sum_{r=1}^{i} a_r$, and consider the distribution*

$$q_i(s) = \frac{a_s}{A_i} \mathbb{1}_{\{s \leq i\}}. \tag{11}$$

Then $\mathcal{A}(S)$ is optimal if and only if

$$\frac{1}{A_t} \geq \sum_{u=t+1}^{n} \frac{1}{(u-1)A_u} \tag{12}$$

for every $t \in [n]$. A sufficient condition is that each weight is at least the average of the preceding weights, that is,

$$a_u \geq \frac{A_{u-1}}{u-1} \tag{13}$$

for every $u \in \{2, \dots, n\}$. In particular, Equation (13) *holds whenever $(a_s)_{s=1}^{n}$ is nondecreasing. This includes the α-power family for every $\alpha \geq 1$.*

Proof. Exact criterion. Because every a_s is positive, $q_t(s) = \frac{a_s}{A_t} > 0$ for every $s \leq t$, so all such pairs (t, s) are feasible. Substituting Equation (11) into Equation (10) gives

$$\frac{a_s}{A_t} \geq a_s \sum_{u=t+1}^{n} \frac{1}{(u-1)A_u}.$$

Cancelling the positive factor a_s yields Equation (12), independently of s. Thus, the criterion is sufficient for optimality of $\mathcal{A}(S)$ by Theorem 10.

For necessity, suppose the criterion fails at some t. Since the sum is empty at $t = n$, this can only happen for $t < n$. Consider a record at time t whose signal has just arrived, that is, the state (t, t). The policy $\mathcal{A}(S)$ has made no earlier selection and accepts this record. This state occurs whenever $I = t$ and $S = t$, so it is reached with positive probability:

$$\mathbb{P}[I = t, S = t] = \frac{1}{n} \frac{a_t}{A_t} > 0.$$

At any such observed record history, stopping succeeds with conditional probability $\frac{t \cdot q_t(t)}{\Psi_{t,t}}$, where

$$t \cdot q_t(t) = \frac{t \cdot a_t}{A_t}.$$

If the policy instead rejects this record and accepts the first later record, its conditional success probability is $\frac{W_{t+1,t}(\{t+1,\ldots,n\})}{\Psi_{t,t}}$, where

$$\begin{aligned} W_{t+1,t}(\{t+1,\ldots,n\}) &= \sum_{u=t+1}^{n} \frac{a_t}{A_u} \prod_{r=t+1}^{u-1} \frac{r-1}{r} \\ &= t \cdot a_t \sum_{u=t+1}^{n} \frac{1}{(u-1)A_u}. \end{aligned}$$

Since Equation (12) fails, the expression for waiting is strictly larger than the expression for stopping. Dividing both by the same positive denominator $\Psi_{t,t}$ shows that rejecting this record gives a strictly higher success probability. Because the state is reached with positive probability, changing the policy only at this state strictly improves its overall success probability. Hence, $\mathcal{A}(S)$ is not optimal, proving necessity.

Sufficient condition on the weights. Assume Equation (13) holds. Dividing that inequality by $A_{u-1}A_u > 0$ gives

$$\frac{1}{(u-1)A_u} \le \frac{a_u}{A_{u-1}A_u} = \frac{1}{A_{u-1}} - \frac{1}{A_u},$$

where the equality uses $a_u = A_u - A_{u-1}$. Summing over $u = t+1, \ldots, n$ cancels all intermediate terms:

$$\sum_{u=t+1}^{n} \frac{1}{(u-1)A_u} \le \sum_{u=t+1}^{n} \left(\frac{1}{A_{u-1}} - \frac{1}{A_u} \right) = \frac{1}{A_t} - \frac{1}{A_n} \le \frac{1}{A_t},$$

which proves Equation (12).

If the weights are nondecreasing, each of $a_1, \ldots, a_{u-1}$ is at most a_u, so

$$A_{u-1} = \sum_{r=1}^{u-1} a_r \le (u-1)a_u,$$

so Equation (13) holds. Finally, every α-power distribution is obtained from $a_s = s^\alpha - (s-1)^\alpha$ and $A_i = i^\alpha$. For $\alpha \ge 1$, the function $x \mapsto x^\alpha$ is convex, so its increases over successive unit intervals are nondecreasing. Thus, these weights satisfy the sufficient condition. □

Corollary 12. *For every signal distribution in our model, the success probability of $\mathcal{A}(S)$, whether or not this policy is optimal, is*

$$\mathbb{P}[\mathcal{A}(S) \textit{ succeeds}] = \frac{1}{n} \left(1 + \sum_{i=2}^{n} \frac{\mathbb{E}[S \mid I = i] - 1}{i-1} \right). \tag{14}$$

Consequently, if $\mathbb{E}[S \mid I = i] \ge 1 + \gamma(i-1)$ for every $i \ge 2$ and some $\gamma \in [0,1]$, then this value is at least $\gamma + \frac{1-\gamma}{n}$. For the nondecreasing-weight family in Equation (11)*, it is at least $\frac{n+1}{2n} > \frac{1}{2}$. This bound is attained when all weights are equal.*

Proof. Exact value. If $I = 1$, then $S = 1$ and $\mathcal{A}(S)$ succeeds with probability 1. Now fix $i \ge 2$ and condition on $I = i$. Let J_i be the position of the best item among the first $i-1$ arrivals. The relative order of these items is uniform, so J_i is uniform on $[i-1]$. The signal is independent of this relative order conditional on $I = i$, so it is also independent of J_i.

Policy $\mathcal{A}(S)$ succeeds exactly when $J_i < S$. Indeed, if $J_i < S$, then the best item among the first $i-1$ arrivals has already appeared before the signal, so there is no record from time S through $i-1$ and the policy accepts the record at i. If $J_i \ge S$, then the item at J_i is a record, so the policy stops

no later than $J_i < i$. For each $s \in [i]$, exactly $s-1$ of the $i-1$ possible positions of J_i lie before s. Using the independence of S and J_i, we obtain

$$\mathbb{P}[\mathcal{A}(S) \text{ succeeds}, S = s \mid I = i] = q_i(s)\mathbb{P}[J_i < s \mid I = i] = q_i(s)\frac{s-1}{i-1}.$$

Summing over s gives

$$\mathbb{P}[\mathcal{A}(S) \text{ succeeds} \mid I = i] = \sum_{s=1}^{i} q_i(s)\frac{s-1}{i-1} = \frac{\mathbb{E}[S \mid I = i] - 1}{i-1}.$$

Finally, averaging over $I \in [n]$, with probability $\frac{1}{n}$ for each position and success probability 1 when $I = 1$, proves Equation (14).

Consequences. Under the assumed bound on $\mathbb{E}[S \mid I = i]$, each of the $n-1$ terms in the sum in Equation (14) is at least γ. Hence,

$$\mathbb{P}[\mathcal{A}(S) \text{ succeeds}] \geq \frac{1 + (n-1)\gamma}{n} = \gamma + \frac{1-\gamma}{n}.$$

For nondecreasing weights, pair position s with $i+1-s$. Since $a_{i+1-s} \geq a_s$ for $s \leq \lfloor i/2 \rfloor$, we have

$$\sum_{s=1}^{i} \left(s - \frac{i+1}{2}\right) a_s = \sum_{s=1}^{\lfloor i/2 \rfloor} \left(\frac{i+1}{2} - s\right)(a_{i+1-s} - a_s) \geq 0.$$

Hence,

$$\mathbb{E}[S \mid I = i] = \frac{1}{A_i}\sum_{s=1}^{i} s \cdot a_s \geq \frac{i+1}{2} = 1 + \frac{i-1}{2}.$$

Taking $\gamma = \frac{1}{2}$ gives the lower bound $\frac{n+1}{2n}$. If all weights are equal, then S is uniform on $[i]$ conditional on $I = i$, so equality holds for every i and the overall value is exactly $\frac{n+1}{2n}$. □

B.2 THE α-POWER FAMILY

We now specialize Theorem 1 to the α-power family. For $s \leq i$, write $q_i(s) = g_s i^{-\alpha}$, where $g_s := s^\alpha - (s-1)^\alpha > 0$. We show that, apart from the factor g_s, the recursion is independent of the observed signal time. Set $f_{n+1} := 0$ and define

$$f_t := \max\left\{f_{t+1}, t^{-\alpha} + \frac{t-1}{t} f_{t+1}\right\} \tag{15}$$

for each $t = n, n-1, \ldots, 1$. We claim that

$$F_{t,s} = g_s f_t \tag{16}$$

for every $1 \leq s \leq t \leq n$. We prove the claim by backward induction on t. For every $s \in [n]$, the initial values satisfy $F_{n+1,s} = g_s f_{n+1} = 0$. Suppose $F_{t+1,s} = g_s f_{t+1}$. Substituting this equality and $q_t(s) = g_s t^{-\alpha}$ into Equation (3) gives

$$\begin{aligned} F_{t,s} &= \max\left\{g_s f_{t+1}, g_s t^{-\alpha} + \frac{t-1}{t} g_s f_{t+1}\right\} \\ &= g_s \max\left\{f_{t+1}, t^{-\alpha} + \frac{t-1}{t} f_{t+1}\right\} \\ &= g_s f_t, \end{aligned}$$

where we used $g_s > 0$ to factor it out of the maximum. This proves Equation (16).

By Theorem 1, stopping at a post-signal record state (t, s) is optimal exactly when $t \cdot q_t(s) \geq F_{t+1,s}$. Using Equation (16), this inequality becomes $g_s t^{1-\alpha} \geq g_s f_{t+1}$. Dividing by the positive factor g_s gives $t^{1-\alpha} \geq f_{t+1}$. Thus, once the signal has arrived, the decision at a record depends only on t, not on the signal time s.

The next lemma evaluates the policy $\mathcal{A}(\max\{S, k\})$, which rejects all items before $\max\{S, k\}$ and then accepts the first record. We will use this formula after determining the optimal thresholds.

Lemma 13. *Fix $\alpha > 0$ and let $k \in [n]$. If $k \geq 2$, the success probability of the threshold policy $\mathcal{A}(\max\{S, k\})$ is*

$$p_{n,\alpha}(k) = \frac{1}{n}\sum_{i=k}^{n}\left(1 - \frac{1}{(i-1)i^{\alpha}}\sum_{r=k}^{i-1} r^{\alpha}\right).$$

For $k = 1$, its success probability is

$$p_{n,\alpha}(1) = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n}\left(1 - \frac{1}{(i-1)i^{\alpha}}\sum_{r=1}^{i-1} r^{\alpha}\right).$$

Proof. Fix $k \in [n]$ and condition on $I = i$.

If $i < k$, the policy rejects the best item because it does not accept before k. If $i = k = 1$, then $S = 1$ and the policy accepts the best item immediately.

Now assume $i \geq \max\{2, k\}$ and additionally condition on $S = s$. The policy accepts the first record at or after $\max\{s, k\}$. It succeeds exactly when the best item among the first $i-1$ arrivals appears before this threshold. Otherwise, that earlier best item is a record at or after the threshold, and the policy stops before i. Conditional on $I = i$, the position of this earlier best item is uniform on $[i-1]$ and independent of S. Thus,

$$\mathbb{P}[\text{success} \mid I = i, S = s] = \frac{\max\{s, k\} - 1}{i - 1}.$$

We average over S and split the sum into $s < k$, where $\max\{s, k\} = k$, and $s \geq k$, where $\max\{s, k\} = s$:

$$\begin{aligned}\mathbb{P}[\text{success} \mid I = i] &= \sum_{s=1}^{i} \frac{s^{\alpha} - (s-1)^{\alpha}}{i^{\alpha}} \cdot \frac{\max\{s, k\} - 1}{i - 1} \\ &= \frac{k-1}{(i-1)i^{\alpha}}\sum_{s=1}^{k-1}\big(s^{\alpha} - (s-1)^{\alpha}\big) + \frac{1}{(i-1)i^{\alpha}}\sum_{s=k}^{i}(s-1)\big(s^{\alpha} - (s-1)^{\alpha}\big).\end{aligned}$$

In the first sum, all intermediate terms cancel, leaving $(k-1)^{\alpha}$. For the second sum, expand the difference and replace $s - 1$ by r in the second part:

$$\begin{aligned}\sum_{s=k}^{i}(s-1)\big(s^{\alpha} - (s-1)^{\alpha}\big) &= \sum_{r=k}^{i}(r-1)r^{\alpha} - \sum_{r=k-1}^{i-1} r^{\alpha+1} \\ &= (i-1)i^{\alpha} - (k-1)^{\alpha+1} - \sum_{r=k}^{i-1} r^{\alpha}.\end{aligned}$$

Substituting both expressions into the average above cancels the terms $(k-1)^{\alpha+1}$ and gives

$$\mathbb{P}[\text{success} \mid I = i] = 1 - \frac{1}{(i-1)i^{\alpha}}\sum_{r=k}^{i-1} r^{\alpha}.$$

Finally, each position $I = i$ has probability $\frac{1}{n}$. If $k \geq 2$, only positions $i \geq k$ can contribute to success, so averaging the expression above gives the first formula. If $k = 1$, the same expression applies to every $i \geq 2$, while the case $I = 1$ contributes an additional $\frac{1}{n}$. This gives the second formula. □

We can now characterize the optimal policies in both regimes, $\alpha \geq 1$ and $0 < \alpha < 1$.

Theorem 2. *For $\alpha \geq 1$, the policy $\mathcal{A}(S)$ is optimal. For $0 < \alpha < 1$, there is a threshold $k_n \in [n]$ such that the policy $\mathcal{A}(\max\{S, k_n\})$ is optimal and $\frac{k_n}{n} \xrightarrow{n\to\infty} (1-\alpha)^{1/\alpha}$. Let $\mathrm{OPT}_n(\alpha)$ denote the optimal success probability on instances of length n. Then*

$$\mathrm{OPT}_n(\alpha) \xrightarrow{n\to\infty} \mathrm{OPT}(\alpha) := \begin{cases} \dfrac{\alpha + (1-\alpha)^{1+1/\alpha}}{\alpha+1}, & 0 < \alpha < 1, \\ \dfrac{\alpha}{\alpha+1}, & \alpha \geq 1\,. \end{cases}$$

Proof. We use the recursion in Equation (15): at a record after the signal, stopping is optimal exactly when $t^{1-\alpha} \geq f_{t+1}$. In each range of α, we first identify an optimal policy and then compute its limiting success probability.

Case 1: $\alpha \geq 1$. Theorem 11 guarantees that $\mathcal{A}(S)$ is optimal. It remains to evaluate its success probability. By Lemma 13,

$$\mathrm{OPT}_n(\alpha) = p_{n,\alpha}(1) = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n}\left(1 - \frac{1}{(i-1)i^\alpha}\sum_{r=1}^{i-1} r^\alpha\right).$$

Since x^α is increasing, the following sum differs from $\int_0^i x^\alpha\,dx$ by at most i^α. Hence, for fixed $\alpha > 0$,

$$\sum_{r=1}^{i-1} r^\alpha = \frac{i^{\alpha+1}}{\alpha+1} + O_\alpha(i^\alpha),$$

and thus

$$1 - \frac{1}{(i-1)i^\alpha}\sum_{r=1}^{i-1} r^\alpha = \frac{\alpha}{\alpha+1} + O_\alpha(i^{-1}).$$

Averaging over i and using $\sum_{i=1}^{n} \frac{1}{i} \leq 1 + \log n$ yields

$$\mathrm{OPT}_n(\alpha) = \frac{\alpha}{\alpha+1} + O_\alpha\left(\frac{\log n}{n}\right),$$

which proves the claimed limit for $\alpha \geq 1$.

Case 2: $0 < \alpha < 1$. To find a candidate threshold, compare accepting a record at t with rejecting it and accepting the first later record. After cancelling the common positive factor $\frac{g_s}{\Psi_{t,s}}$ from their success probabilities, the first expression minus the second is G_t, defined for $t \in [n]$ by

$$G_t := t^{1-\alpha} - t\sum_{u=t+1}^{n}\frac{1}{(u-1)u^\alpha}, \qquad k_n := \min\{t \in [n] : G_t \geq 0\}.$$

The index k_n is well-defined because the sum defining G_n is empty and hence $G_n = n^{1-\alpha} > 0$. Moreover, for every $1 \leq t < n$,

$$\begin{aligned}\frac{G_{t+1}}{t+1} - \frac{G_t}{t} &= (t+1)^{-\alpha} - t^{-\alpha} + \frac{1}{t(t+1)^\alpha} \\ &= \frac{(t+1)^{1-\alpha} - t^{1-\alpha}}{t} > 0.\end{aligned}$$

Thus, $\frac{G_t}{t}$ is strictly increasing. Since dividing by $t > 0$ does not change the sign, G_t can change from negative to nonnegative at most once. By the definition of k_n, we have $G_t \geq 0$ for every $t \geq k_n$ and $G_t < 0$ for every $t < k_n$.

We now verify that this candidate threshold satisfies the optimal stopping rule, choosing to stop when both actions have equal value. First work backward from $t = n$ to $t = k_n$. Suppose accepting every later record is optimal. Similar to the proof of Theorem 11 the stopping set $\{t+1, \ldots, n\}$ then gives

$$f_{t+1} = t\sum_{u=t+1}^{n}\frac{1}{(u-1)u^\alpha}.$$

Since $t \geq k_n$, we have $G_t \geq 0$, which is exactly $t^{1-\alpha} \geq f_{t+1}$. Therefore, stopping is optimal at t. This proves that accepting every record at or after k_n is optimal once the signal has arrived.

If $k_n = 1$, this already proves that $\mathcal{A}(S) = \mathcal{A}(\max\{S, k_n\})$ is optimal. Suppose, therefore, that $k_n > 1$. Since accepting every record from k_n onward is optimal after the signal, the fixed-set formula gives

$$f_{k_n} = \sum_{u=k_n}^{n} u^{-\alpha}\prod_{r=k_n}^{u-1}\left(1 - \frac{1}{r}\right)$$

$$= (k_n - 1) \sum_{u=k_n}^{n} \frac{1}{(u-1)u^\alpha}.$$

Write $B_n := f_{k_n}$ for this value. Since $G_{k_n-1} = (k_n-1)^{1-\alpha} - B_n < 0$, we have

$$B_n > (k_n - 1)^{1-\alpha}.$$

We now show that rejecting every record before k_n is optimal and that f_t stays equal to B_n throughout these earlier times. At $t = k_n - 1$, we have $f_{t+1} = B_n > t^{1-\alpha}$, so rejecting is strictly better and the recursion gives $f_t = f_{t+1} = B_n$. Inductively, if $t < k_n - 1$ and $f_{t+1} = B_n$, then, since $1 - \alpha > 0$,

$$t^{1-\alpha} \le (k_n - 1)^{1-\alpha} < B_n = f_{t+1}.$$

Thus, rejecting is again strictly better and $f_t = B_n$. Since stopping before the signal cannot succeed either, an optimal rule rejects all items before $\max\{S, k_n\}$ and then accepts the first record. This is exactly $\mathcal{A}(\max\{S, k_n\})$.

Limit of the threshold. Fix $c \in (0, 1]$ and let $t_n = \lfloor cn \rfloor$. For sufficiently large n, we have $t_n \ge 1$, and

$$\begin{aligned} n^{\alpha-1} G_{t_n} &= \left(\frac{t_n}{n}\right)^{1-\alpha} - \frac{t_n}{n} \sum_{u=t_n+1}^{n} \frac{n^\alpha}{(u-1)u^\alpha} \\ &= \left(\frac{t_n}{n}\right)^{1-\alpha} - \frac{t_n}{n} \cdot \frac{1}{n} \sum_{u=t_n+1}^{n} \left(\frac{u}{n}\right)^{-(\alpha+1)} \frac{u}{u-1}. \end{aligned}$$

For all sufficiently large n, every index in the sum satisfies $\frac{u}{n} \ge \frac{c}{2}$ and $0 \le \frac{u}{u-1} - 1 \le \frac{1}{t_n}$. Replacing the factor $\frac{u}{u-1}$ by 1 therefore changes the average by at most $\frac{(c/2)^{-(\alpha+1)}}{t_n}$, which tends to zero as n goes to infinity. The remaining average is a right Riemann sum for $y^{-(\alpha+1)}$ on $\left[\frac{t_n}{n}, 1\right]$. Its lower endpoint differs from c by at most $\frac{1}{n}$, and the integrand is continuous and bounded on $\left[\frac{c}{2}, 1\right]$. Hence,

$$\begin{aligned} n^{\alpha-1} G_{t_n} &\longrightarrow c^{1-\alpha} - c \int_c^1 y^{-(\alpha+1)}\, dy \\ &= c^{1-\alpha} - c\frac{c^{-\alpha} - 1}{\alpha} = \frac{c - (1-\alpha)c^{1-\alpha}}{\alpha}. \end{aligned}$$

The limiting expression is negative for $c < (1-\alpha)^{1/\alpha}$ and positive for $c > (1-\alpha)^{1/\alpha}$. Choose $0 < c_- < (1-\alpha)^{1/\alpha} < c_+ < 1$. For all sufficiently large n, the preceding convergence gives $G_{\lfloor c_- n \rfloor} < 0 < G_{\lfloor c_+ n \rfloor}$. Since $\frac{G_t}{t}$ is strictly increasing and k_n is the first index with $G_{k_n} \ge 0$, we obtain

$$\lfloor c_- n \rfloor < k_n \le \lfloor c_+ n \rfloor, \qquad \text{so} \qquad c_- < \frac{k_n}{n} \le c_+.$$

Choosing c_- and c_+ arbitrarily close to $(1-\alpha)^{1/\alpha}$ proves that

$$\frac{k_n}{n} \xrightarrow{n\to\infty} (1-\alpha)^{1/\alpha} =: c_\alpha.$$

Limiting success probability. Since $\frac{k_n}{n} \to c_\alpha > 0$, we have $k_n \ge 2$ for all sufficiently large n. Applying Lemma 13 with $k = k_n$ gives

$$\mathrm{OPT}_n(\alpha) = p_{n,\alpha}(k_n) = \frac{1}{n} \sum_{i=k_n}^{n} \left(1 - \frac{1}{(i-1)i^\alpha} \sum_{r=k_n}^{i-1} r^\alpha\right)$$

Set $c_n := \frac{k_n}{n}$ and $x_i := \frac{i}{n}$. We first replace the inner sum by an integral, keeping an error bound that applies to every term of the outer average. Since $c_n \to c_\alpha > 0$, for all sufficiently large n and every $i \ge k_n$, both x_i and $x_i - \frac{1}{n}$ are at least $\frac{c_\alpha}{2}$. We have

$$\frac{1}{(i-1)i^\alpha} \sum_{r=k_n}^{i-1} r^\alpha = \frac{1}{(x_i - 1/n)x_i^\alpha} \left(\frac{1}{n} \sum_{r=k_n}^{i-1} \left(\frac{r}{n}\right)^\alpha\right).$$

Since z^α is increasing, the integral lies between the sums of rectangles using the left and right endpoints of each interval. Their difference gives

$$0 \le \int_{c_n}^{x_i} z^\alpha \, dz - \frac{1}{n} \sum_{r=k_n}^{i-1} \left(\frac{r}{n}\right)^\alpha \le \frac{x_i^\alpha - c_n^\alpha}{n} \le \frac{1}{n}.$$

Also,

$$\frac{1}{(x_i - 1/n)x_i^\alpha} - \frac{1}{x_i^{\alpha+1}} = \frac{1}{n(x_i - 1/n)x_i^{\alpha+1}}.$$

The denominators are bounded below, and the inner sums and integrals are at most 1. Thus, replacing the inner sum by its integral and $x_i - \frac{1}{n}$ by x_i changes each term of the outer average by at most a constant depending only on α times $\frac{1}{n}$. Averaging cannot increase this error bound.

Next, replacing c_n by c_α inside the integral changes each term by at most $(2/c_\alpha)^{\alpha+1}|c_n - c_\alpha|$, which tends to zero. Replacing the lower summation index k_n by $\lceil c_\alpha n \rceil$ adds or removes at most $n|c_n - c_\alpha| + 1$ terms. These terms are bounded, so their contribution after division by n also tends to zero. What remains is an ordinary Riemann sum for a continuous function on $[c_\alpha, 1]$. Therefore,

$$\frac{1}{n} \sum_{i=k_n}^{n} \left(1 - \frac{1}{(i-1)i^\alpha} \sum_{r=k_n}^{i-1} r^\alpha\right) \longrightarrow \int_{c_\alpha}^{1} \left(1 - \frac{1}{y^{\alpha+1}} \int_{c_\alpha}^{y} z^\alpha \, dz\right) dy.$$

Evaluating the inner integral gives $\int_{c_\alpha}^{y} z^\alpha \, dz = \frac{y^{\alpha+1} - c_\alpha^{\alpha+1}}{\alpha+1}$. Thus, the integrand becomes

$$1 - \frac{1}{y^{\alpha+1}} \frac{y^{\alpha+1} - c_\alpha^{\alpha+1}}{\alpha+1} = \frac{\alpha}{\alpha+1} + \frac{c_\alpha^{\alpha+1}}{(\alpha+1)y^{\alpha+1}}.$$

The outer integral therefore evaluates to

$$\begin{aligned} \frac{\alpha}{\alpha+1}(1 - c_\alpha) + \frac{c_\alpha - c_\alpha^{\alpha+1}}{\alpha(\alpha+1)} &= \frac{\alpha^2 + (1-\alpha^2)c_\alpha - c_\alpha^{\alpha+1}}{\alpha(\alpha+1)} \\ &= \frac{\alpha^2 + \alpha(1-\alpha)c_\alpha}{\alpha(\alpha+1)} = \frac{\alpha + (1-\alpha)^{1+1/\alpha}}{\alpha+1}. \end{aligned}$$

Here we used $c_\alpha^{\alpha+1} = (1-\alpha)c_\alpha$ to simplify the second line. This proves the claimed limit for $0 < \alpha < 1$. □

C Omitted proofs from Section 3

We first compute the limiting success probability for any fixed threshold fraction. We then identify the best fraction and analyze what happens when it is chosen using an estimated parameter.

Lemma 14. *Fix $\alpha > 0$ and $\beta \in [0, 1]$. Recall that $\widetilde{k}_n(\beta) = \lceil \beta n \rceil$ and that $p_{n,\alpha}(k)$ denotes the success probability of $\mathcal{A}(\max\{S, k\})$. Then*

$$p_{n,\alpha}(\widetilde{k}_n(\beta)) \xrightarrow{n\to\infty} f(\alpha, \beta) := \frac{\alpha}{\alpha+1} + \frac{1-\alpha}{\alpha}\beta - \frac{\beta^{\alpha+1}}{\alpha(\alpha+1)}.$$

Proof. If $\beta = 0$, then $\widetilde{k}_n(\beta) = 0$ and the policy is $\mathcal{A}(S)$. The claim therefore follows from Equation (2). If $\beta = 1$, then $\widetilde{k}_n(\beta) = n$ and the policy succeeds exactly when $I = n$. Hence, its success probability is $\frac{1}{n} \to 0 = f(\alpha, 1)$.

Now suppose $0 < \beta < 1$ and set $\ell_n := \widetilde{k}_n(\beta) = \lceil \beta n \rceil$. For all sufficiently large n, we have $\ell_n \ge 2$, so Lemma 13 gives

$$p_{n,\alpha}(\ell_n) = \frac{1}{n} \sum_{i=\ell_n}^{n} \left(1 - \frac{1}{(i-1)i^\alpha} \sum_{r=\ell_n}^{i-1} r^\alpha\right).$$

Set $c_n := \frac{\ell_n}{n}$ and $x_i := \frac{i}{n}$. As in the proof of Theorem 2, we replace each inner sum by an integral. Since z^α is increasing, comparison with rectangles gives

$$0 \le \int_{c_n}^{x_i} z^\alpha \, dz - \frac{1}{n} \sum_{r=\ell_n}^{i-1} \left(\frac{r}{n}\right)^\alpha \le \frac{x_i^\alpha - c_n^\alpha}{n} \le \frac{1}{n}.$$

Also, $0 \le c_n - \beta < \frac{1}{n}$, so replacing the lower integration limit by β changes the integral by at most $\frac{1}{n}$. For all sufficiently large n, both x_i and $x_i - \frac{1}{n}$ are at least $\frac{\beta}{2}$. Thus, replacing $x_i - \frac{1}{n}$ by x_i in the denominator below also changes the expression by at most a constant times $\frac{1}{n}$. We obtain

$$\begin{aligned}\frac{1}{(i-1)i^\alpha} \sum_{r=\ell_n}^{i-1} r^\alpha &= \frac{1}{(x_i - 1/n)x_i^\alpha} \left(\frac{1}{n} \sum_{r=\ell_n}^{i-1} \left(\frac{r}{n}\right)^\alpha\right) \\ &= \frac{1}{x_i^{\alpha+1}} \int_\beta^{x_i} z^\alpha \, dz + O_{\alpha,\beta}\Big(\frac{1}{n}\Big),\end{aligned}$$

with the same error bound for every $i \ge \ell_n$. Averaging does not increase this bound. The remaining outer sum is a Riemann sum for a continuous function on $[\beta, 1]$, so

$$p_{n,\alpha}(\ell_n) \longrightarrow \int_\beta^1 \left(1 - \frac{1}{y^{\alpha+1}} \int_\beta^y z^\alpha \, dz\right) dy.$$

The integrand simplifies to

$$1 - \frac{1}{y^{\alpha+1}} \int_\beta^y z^\alpha \, dz = \frac{\alpha}{\alpha+1} + \frac{\beta^{\alpha+1}}{(\alpha+1)y^{\alpha+1}}.$$

Integrating these two terms gives

$$\begin{aligned}\int_\beta^1 \left(\frac{\alpha}{\alpha+1} + \frac{\beta^{\alpha+1}}{(\alpha+1)y^{\alpha+1}}\right) dy &= \frac{\alpha}{\alpha+1}(1-\beta) + \frac{\beta^{\alpha+1}}{\alpha+1} \int_\beta^1 y^{-\alpha-1} \, dy \\ &= \frac{\alpha}{\alpha+1}(1-\beta) + \frac{\beta^{\alpha+1}}{\alpha+1} \frac{\beta^{-\alpha} - 1}{\alpha} \\ &= \frac{\alpha}{\alpha+1} + \frac{1-\alpha}{\alpha}\beta - \frac{\beta^{\alpha+1}}{\alpha(\alpha+1)} = f(\alpha, \beta).\end{aligned}$$

□

We next verify the optimal threshold fraction used in the main text.

Lemma 15. *For every fixed $\alpha > 0$, the function $f(\alpha, \cdot)$ is maximized over $[0, 1]$ by*

$$\beta^\star(\alpha) = \begin{cases} (1-\alpha)^{1/\alpha}, & 0 < \alpha < 1, \\ 0, & \alpha \ge 1. \end{cases}$$

Moreover, $f(\alpha, \beta^\star(\alpha)) = \mathrm{OPT}(\alpha)$.

Proof. Differentiating the expression in Lemma 14 gives

$$\partial_\beta f(\alpha, \beta) = \frac{1 - \alpha - \beta^\alpha}{\alpha} \qquad \text{for } \beta \in (0, 1).$$

If $0 < \alpha < 1$, this derivative is positive when $\beta < (1-\alpha)^{1/\alpha}$ and negative when $\beta > (1-\alpha)^{1/\alpha}$. Thus, $f(\alpha, \cdot)$ increases up to this fraction and decreases afterwards, so its unique maximizer is $\beta^\star(\alpha) = (1-\alpha)^{1/\alpha}$. If $\alpha \ge 1$, the derivative is negative throughout $(0, 1)$, so the maximum is attained at $\beta = 0$.

For $0 < \alpha < 1$, the identity $(\beta^\star(\alpha))^\alpha = 1 - \alpha$ gives

$$f(\alpha, \beta^\star(\alpha)) = \frac{\alpha + (1-\alpha)\beta^\star(\alpha)}{\alpha+1} = \frac{\alpha + (1-\alpha)^{1+1/\alpha}}{\alpha+1}.$$

For $\alpha \ge 1$, we have $f(\alpha, 0) = \frac{\alpha}{\alpha+1}$. In both cases, the result equals $\mathrm{OPT}(\alpha)$ from Theorem 2. □

We can now prove Theorem 3 for an estimated parameter $\hat{\alpha}$.

Theorem 3. *Let $\alpha, \hat{\alpha} > 0$, where the true signal is α-power and the threshold is chosen using $\hat{\alpha}$. Set $\widehat{k}_n := \widetilde{k}_n(\beta^\star(\hat{\alpha})) = \lceil \beta^\star(\hat{\alpha}) n \rceil$. Then the policy $\mathcal{A}(\max\{S, \widehat{k}_n\})$ satisfies $\lim_{n\to\infty} \mathbb{P}[\textit{success}] = g(\alpha, \hat{\alpha}) := f(\alpha, \beta^\star(\hat{\alpha}))$. Explicitly,*

$$g(\alpha, \hat{\alpha}) = \begin{cases} \dfrac{\alpha}{\alpha+1} + \dfrac{1-\alpha}{\alpha}(1-\hat{\alpha})^{1/\hat{\alpha}} - \dfrac{(1-\hat{\alpha})^{(\alpha+1)/\hat{\alpha}}}{\alpha(\alpha+1)}, & 0 < \hat{\alpha} < 1, \\ \dfrac{\alpha}{\alpha+1}, & \hat{\alpha} \geq 1. \end{cases}$$

In particular, if $\hat{\alpha} \leq \alpha$, then $g(\alpha, \hat{\alpha}) \geq \mathrm{OPT}(\hat{\alpha}) \geq 1/e$. Moreover, for $0 < \alpha, \hat{\alpha} < 1$, let $D(\alpha, \hat{\alpha}) := \mathrm{OPT}(\alpha) - g(\alpha, \hat{\alpha})$. As $\delta \to 0$ subject to $0 < \alpha + \delta < 1$, we have $D(\alpha, \alpha+\delta) = C_\alpha \delta^2 + O_\alpha(|\delta|^3)$ for some constant $C_\alpha > 0$.

Proof. We first compute the success probability under an estimated parameter, then prove the conservative guarantee and the quadratic loss for small estimation errors.

Success probability under an estimated parameter. The policy uses the fixed fraction $\beta^\star(\hat{\alpha})$, while the signal has true parameter α. By Lemma 14, its limiting success probability is therefore $g(\alpha, \hat{\alpha}) = f(\alpha, \beta^\star(\hat{\alpha}))$. If $0 < \hat{\alpha} < 1$, substituting $\beta^\star(\hat{\alpha}) = (1-\hat{\alpha})^{1/\hat{\alpha}}$ gives the first branch of the stated formula. If $\hat{\alpha} \geq 1$, then $\beta^\star(\hat{\alpha}) = 0$, giving $g(\alpha, \hat{\alpha}) = \frac{\alpha}{\alpha+1}$.

Conservative estimates. Assume that $\hat{\alpha} \leq \alpha$. If $\hat{\alpha} \geq 1$, then

$$g(\alpha, \hat{\alpha}) = \frac{\alpha}{\alpha+1} \geq \frac{\hat{\alpha}}{\hat{\alpha}+1} = \mathrm{OPT}(\hat{\alpha}),$$

because $x \mapsto \frac{x}{x+1}$ is increasing.

Now suppose $0 < \hat{\alpha} < 1$, and set $\beta := \beta^\star(\hat{\alpha}) \in (0,1)$. We keep this threshold fraction fixed and show that its success probability increases with the true parameter. Rewrite the expression in Lemma 14 as

$$\begin{aligned} f(\alpha, \beta) &= 1 - \beta - \frac{1}{\alpha}\left(1 - \beta - \frac{1-\beta^{\alpha+1}}{\alpha+1}\right) \\ &= 1 - \beta - \frac{1}{\alpha}\int_\beta^1 (1 - t^\alpha)\, dt. \end{aligned}$$

For every $t \in (0,1)$,

$$\frac{1-t^\alpha}{\alpha} = \int_t^1 u^{\alpha-1}\, du$$

is decreasing in α: for $0 < u < 1$, increasing the exponent decreases $u^{\alpha-1}$. Since this integral is subtracted in the expression for $f(\cdot,\cdot)$, the success probability $f(\alpha, \beta)$ increases with α. Using $\alpha \geq \hat{\alpha}$ and Lemma 15, we obtain

$$g(\alpha, \hat{\alpha}) = f(\alpha, \beta^\star(\hat{\alpha})) \geq f(\hat{\alpha}, \beta^\star(\hat{\alpha})) = \mathrm{OPT}(\hat{\alpha}).$$

It remains to prove $\mathrm{OPT}(\hat{\alpha}) \geq 1/e$. For every deterministic threshold k, the policy $\mathcal{A}(\max\{S, k\})$ succeeds on every realization where the classic policy $\mathcal{A}(k)$ succeeds. Indeed, on such a realization, $k \leq I$ and there is no record from time k through $I-1$. Since $S \leq I$, the later threshold satisfies $k \leq \max\{S, k\} \leq I$. There is still no record between this threshold and $I-1$, so $\mathcal{A}(\max\{S,k\})$ also accepts the best item at I. In particular, taking the classic threshold $k = \max\{1, \lceil n/e \rceil\}$ gives a success probability at least that of the classic policy, which tends to $1/e$. The optimal success probability $\mathrm{OPT}_n(\hat{\alpha})$ is at least as large. Since $\mathrm{OPT}_n(\hat{\alpha}) \to \mathrm{OPT}(\hat{\alpha})$ by Theorem 2, the benchmark follows.

Local quadratic loss. Fix the true parameter $0 < \alpha < 1$ and vary only the estimated parameter a in a neighborhood of α contained in $(0,1)$. Write

$$b(a) := \beta^\star(a) = (1-a)^{1/a}, \qquad h(a) := f(\alpha, b(a)).$$

Thus, $h(a)$ is the limiting success probability when the estimate is a. By Lemma 15, $h(\alpha) = \mathrm{OPT}(\alpha)$, so

$$D(\alpha, \alpha+\delta) = h(\alpha) - h(\alpha+\delta).$$

The derivative of $f(\alpha, \cdot)$ vanishes at its interior maximizer $b(\alpha)$. Differentiating the formula for $f(\cdot, \cdot)$ with respect to β twice also gives its second derivative there:

$$\partial_\beta f(\alpha, b(\alpha)) = 0, \qquad \partial^2_{\beta\beta} f(\alpha, b(\alpha)) = -b(\alpha)^{\alpha-1}.$$

To differentiate b, use $\log b(a) = \frac{\log(1-a)}{a}$. This gives

$$\frac{b'(a)}{b(a)} = -\frac{1}{a(1-a)} - \frac{\log(1-a)}{a^2},$$

and hence

$$b'(\alpha) = -\frac{b(\alpha)}{\alpha^2}\left(\frac{\alpha}{1-\alpha} + \log(1-\alpha)\right).$$

By the chain rule,

$$h'(a) = \partial_\beta f(\alpha, b(a)) b'(a),$$
$$h''(a) = \partial^2_{\beta\beta} f(\alpha, b(a)) b'(a)^2 + \partial_\beta f(\alpha, b(a)) b''(a).$$

At $a = \alpha$, the factors $\partial_\beta f(\alpha, b(\alpha))$ vanish. Consequently,

$$h'(\alpha) = 0, \qquad h''(\alpha) = -b(\alpha)^{\alpha-1} b'(\alpha)^2.$$

In particular, there is no linear term in the loss. Since h has a continuous third derivative near α, Taylor's series expansion gives

$$D(\alpha, \alpha+\delta) = -\frac{1}{2} h''(\alpha)\delta^2 + O_\alpha(|\delta|^3) = C_\alpha \delta^2 + O_\alpha(|\delta|^3),$$

where

$$C_\alpha = \frac{1}{2} b(\alpha)^{\alpha-1} b'(\alpha)^2 = \frac{b(\alpha)^{\alpha+1}}{2\alpha^4}\left(\frac{\alpha}{1-\alpha} + \log(1-\alpha)\right)^2 > 0.$$

To see that the squared factor is nonzero, note that

$$-\log(1-\alpha) = \int_0^\alpha \frac{1}{1-x} dx < \frac{\alpha}{1-\alpha},$$

since the integrand is strictly smaller than $\frac{1}{1-\alpha}$ for $x < \alpha$. Together with $b(\alpha) > 0$, this proves $C_\alpha > 0$. □

C.1 Learning a conservative parameter from historical episodes

The robustness guarantee assumes that the parameter estimate is conservative. We now obtain such an estimate from completed historical episodes and connect it to the success probability of the resulting policy. The guarantee holds for finite training samples and finite test instances. It also gives convergence to the optimal success probability as the number of training episodes grows.

Training data and estimator. Suppose independent historical episodes reveal the best-item position I_r and signal time S_r after completion. Their lengths may differ, but all signals follow the same conditional α-power distribution, for an unknown $\alpha > 0$. Episodes with $I_r = 1$ always have $S_r = 1$ and contain no information about α, so we discard them. Relabel the remaining pairs as $\mathcal{D} = ((I_r, S_r))_{r=1}^m$. If $m = 0$, set $\alpha_{\mathrm{L}} = 0$ and use $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$, which uses the classic threshold and requires no estimate of α. Otherwise, we condition on the observed positions $\mathcal{I} = (I_1, \ldots, I_m)$ and treat them as fixed. The randomness in the estimate then comes only from the signals; we do not need to estimate the distribution of the positions.

For the retained episodes, define:

- $t_r := \frac{\lfloor I_r/2 \rfloor}{I_r}$, the cutoff as a fraction of the best-item position;

- $Y_r := \mathbb{1}\{S_r \le \lfloor I_r/2 \rfloor\}$, indicating whether the signal arrived by this cutoff;
- $\overline{Y} := \frac{1}{m}\sum_{r=1}^{m} Y_r$, the observed fraction of signals arriving by their cutoffs.

The α-power distribution gives $\mathbb{P}[Y_r = 1 \mid \mathcal{I}] = t_r^\alpha$. Thus the conditional mean of $\overline{Y}$ is the known function

$$h_{\mathcal{I}}(a) := \frac{1}{m}\sum_{r=1}^{m} t_r^a$$

evaluated at the unknown parameter α. Since $\frac{1}{3} \le t_r \le \frac{1}{2}$, this function decreases continuously and strictly from 1 to 0 as a increases from 0 to infinity. Using the actual ratios t_r accounts for integer rounding, including odd best-item positions.

Larger values of α make early signals less likely, so an upper bound on this mean gives a lower bound on α. Fix a desired failure probability $\delta \in (0, 1)$ and define:

- the error margin $\varepsilon_m := \sqrt{\frac{\log(2/\delta)}{2m}}$;
- the upper confidence bound $u := \min\{1, \overline{Y} + \varepsilon_m\}$ for the mean;
- the conservative parameter estimate

$$\alpha_{\mathrm{L}} := h_{\mathcal{I}}^{-1}(u). \tag{17}$$

Thus, α_{L} is the unique nonnegative solution of $h_{\mathcal{I}}(a) = u$. This solution exists because $0 < u \le 1$, and can be found by bisection. In particular, $u = 1$ gives $\alpha_{\mathrm{L}} = 0$. If all retained positions are even, then $h_{\mathcal{I}}(a) = 2^{-a}$ and $\alpha_{\mathrm{L}} = -\log_2 u$.

Policy and guarantee. For $a > 0$, let $k_n(a)$ be an exact optimal threshold from Theorem 2, choosing $k_n(a) = 1$ for $a \ge 1$. Set $k_n(0) = k_n^{\mathrm{cl}}$ and $\mathrm{OPT}_n(0) = p_n^{\mathrm{cl}}(k_n^{\mathrm{cl}})$. After training, use

$$\pi_{\mathcal{D},n} := \mathcal{A}(\max\{S, k_n(\alpha_{\mathrm{L}})\})$$

on a fresh, independent random-order test episode of length n with the same parameter α. The test signal must be independent of relative ranks conditional on the best-item position. Write $P_{\mathcal{D},n}(\alpha)$ for the success probability on this test episode, conditional on the training data. This is the policy's true success probability, not a success fraction computed from finitely many test runs.

Theorem 16. *Fix any vector of retained training positions $\mathcal{I}$ with $m \ge 1$. With probability at least $1-\delta$ over the training signals, conditional on these positions, the estimator in Equation* (17) *satisfies*

$$0 \le \alpha_{\mathrm{L}} \le \alpha, \qquad \alpha - \alpha_{\mathrm{L}} \le \frac{2 \cdot 3^\alpha}{\log 2}\,\varepsilon_m. \tag{18}$$

On the same event, simultaneously for every test size $n \ge 1$,

$$P_{\mathcal{D},n}(\alpha) \ge \mathrm{OPT}_n(\alpha_{\mathrm{L}}) \ge p_n^{\mathrm{cl}}(k_n^{\mathrm{cl}}), \tag{19}$$

$$0 \le \mathrm{OPT}_n(\alpha) - P_{\mathcal{D},n}(\alpha) \le \min\left\{1, \frac{2 \cdot 3^\alpha}{\log 2}\,\varepsilon_m\right\}. \tag{20}$$

If $\alpha_{\mathrm{L}} \ge 1$, the learned policy is exactly optimal for every n on this event. If $m = 0$, the policy $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$ has success probability at least the finite-n classic benchmark for every $\alpha > 0$.

Proof. We first bound the error in the observed fraction of early signals, then translate this into a bound on the estimated parameter. We next prove the classic benchmark and bound the loss relative to the true optimum.

A confidence bound for the mean. Conditional on $\mathcal{I}$, the variables $Y_1, \ldots, Y_m$ are independent, take values in $[0, 1]$, and have average mean $h_{\mathcal{I}}(\alpha)$. Hoeffding's inequality gives

$$\mathbb{P}\left[\left|\overline{Y} - h_{\mathcal{I}}(\alpha)\right| > \varepsilon_m \;\middle|\; \mathcal{I}\right] \le 2e^{-2m\varepsilon_m^2} = \delta.$$

For the rest of the proof, suppose that $|\overline{Y} - h_{\mathcal{I}}(\alpha)| \le \varepsilon_m$, which holds with probability at least $1-\delta$. Adding ε_m to $\overline{Y}$ gives a value between $h_{\mathcal{I}}(\alpha)$ and $h_{\mathcal{I}}(\alpha) + 2\varepsilon_m$. Replacing this value by 1 if it exceeds 1 cannot put it below $h_{\mathcal{I}}(\alpha)$, since this mean is at most 1. Therefore,

$$h_{\mathcal{I}}(\alpha) \le u \le h_{\mathcal{I}}(\alpha) + 2\varepsilon_m. \tag{21}$$

A conservative parameter estimate. By definition, $h_{\mathcal{I}}(\alpha_{\mathrm{L}}) = u$. Since $h_{\mathcal{I}}$ is decreasing and $u \ge h_{\mathcal{I}}(\alpha)$, we must have $\alpha_{\mathrm{L}} \le \alpha$. To bound how much smaller the estimate can be, we bound the rate at which $h_{\mathcal{I}}$ decreases. For $0 \le a \le \alpha$, differentiation gives

$$-h'_{\mathcal{I}}(a) = \frac{1}{m}\sum_{r=1}^{m} t_r^a \log(1/t_r) \ge 3^{-\alpha}\log 2.$$

Indeed, $t_r \ge \frac{1}{3}$ and $a \le \alpha$ give $t_r^a \ge 3^{-\alpha}$, while $t_r \le \frac{1}{2}$ gives $\log(1/t_r) \ge \log 2$. Integrating from α_{L} to α now yields

$$\begin{aligned} u - h_{\mathcal{I}}(\alpha) &= h_{\mathcal{I}}(\alpha_{\mathrm{L}}) - h_{\mathcal{I}}(\alpha) \\ &= \int_{\alpha_{\mathrm{L}}}^{\alpha} -h'_{\mathcal{I}}(a)\, da \\ &\ge 3^{-\alpha}\log 2\,(\alpha - \alpha_{\mathrm{L}}). \end{aligned}$$

The left-hand side is at most $2\varepsilon_m$ by Equation (21). Dividing by $3^{-\alpha}\log 2 > 0$ gives the stated bound on $\alpha - \alpha_{\mathrm{L}}$. This proves Equation (18), including the case $\alpha_{\mathrm{L}} = 0$.

The classic benchmark. Fix any test size n and a parameter $0 < a \le \alpha$. We compare the two signal distributions on the same random permutation. Draw a single number U uniformly from $(0,1)$, independently of this permutation, and define:

- $S_a := \lceil IU^{1/a} \rceil$, the signal for parameter a;
- $S_\alpha := \lceil IU^{1/\alpha} \rceil$, the signal for the true parameter α.

For every $s \in [i]$,

$$\mathbb{P}[S_a \le s \mid I = i] = \mathbb{P}[U \le (s/i)^a] = (s/i)^a,$$

and the same calculation applies to S_α. Thus, each signal has the required conditional α-power distribution. Moreover, $a \le \alpha$ gives $U^{1/a} \le U^{1/\alpha}$, so $S_a \le S_\alpha \le I$ on every realization.

For any fixed threshold k, suppose that $\mathcal{A}(\max\{S_a, k\})$ selects the best item. Then $k \le I$ and there is no record from time $\max\{S_a, k\}$ through $I-1$. The later threshold $\max\{S_\alpha, k\}$ still lies at or before I. There is also no record from this later threshold through $I-1$, so the policy with signal S_α accepts the same best item. Its success probability is therefore at least as large.

If $\alpha_{\mathrm{L}} > 0$, take $a = \alpha_{\mathrm{L}}$ and $k = k_n(a)$. The policy is optimal under parameter a, so the comparison gives

$$P_{\mathcal{D},n}(\alpha) \ge \mathrm{OPT}_n(\alpha_{\mathrm{L}}).$$

The optimum on the right is at least the classic benchmark because a policy can ignore the signal. If $\alpha_{\mathrm{L}} = 0$, apply the same record argument to the classic threshold: changing k_n^{cl} to $\max\{S, k_n^{\mathrm{cl}}\} \le I$ on a classic success preserves that success. This proves Equation (19) and also covers the case $m = 0$. The only restriction on the training data used here is $\alpha_{\mathrm{L}} \le \alpha$. Thus, the same event of probability at least $1-\delta$ gives the guarantee for every n.

Loss relative to the true optimum. We first show that increasing the parameter by $b-a$ can increase the optimal success probability by at most $b-a$. Recall that $p_{n,a}(k)$ is the success probability of $\mathcal{A}(\max\{S,k\})$ when the signal has parameter a. For a fixed threshold $k \in [n]$, Lemma 13 gives

$$p_{n,a}(k) = \frac{1}{n}\left[\mathbb{1}\{k=1\} + \sum_{i=\max\{2,k\}}^{n}\left(1 - \frac{1}{i-1}\sum_{s=k}^{i-1}\left(\frac{s}{i}\right)^a\right)\right].$$

Here the outer average is over the uniformly distributed best-item position I. Positions $i < k$ give no success, and the indicator accounts for $i = 1$, where success is possible exactly when $k = 1$. The

formula also holds at $a = 0$ if we put the signal at time 1; the resulting policy is the classic threshold policy. Differentiating the finite sums, for $a \geq 0$ we obtain

$$0 \leq \frac{\partial}{\partial a} p_{n,a}(k) = \frac{1}{n} \sum_{i=\max\{2,k\}}^{n} \frac{1}{i-1} \sum_{s=k}^{i-1} \left(\frac{s}{i}\right)^a \log\left(\frac{i}{s}\right).$$

In order to give an upper bound as well, we first use $\left(\frac{s}{i}\right)^a \leq 1$ and replace the lower limit of the inner sum by $s = 1$. Then, we write $\log\left(\frac{s}{i}\right) = \sum_{j=s}^{i-1} \log\left(1 + \frac{1}{j}\right)$ to obtain

$$\sum_{s=k}^{i-1} \left(\frac{s}{i}\right)^a \log\left(\frac{i}{s}\right) \leq \sum_{s=1}^{i-1} \log\left(\frac{i}{s}\right) \leq \sum_{s=1}^{i-1} \left(\sum_{j=s}^{i-1} \log\left(1 + \frac{1}{j}\right) \right).$$

Now for each fixed j, the term $\log\left(1 + \frac{1}{j}\right)$ appears for exactly j choices of s. Hence

$$\begin{aligned} &= \sum_{j=1}^{i-1} j \log\left(1 + \frac{1}{j}\right) \\ &\leq \sum_{j=1}^{i-1} j \cdot \frac{1}{j} = i - 1, \end{aligned}$$

using $\log(1 + x) \leq x$ for $x \geq 0$ in the last inequality. Each inner sum in the derivative is therefore at most $i - 1$. Combining, we obtain

$$\frac{1}{n} \sum_{i=\max\{2,k\}}^{n} \frac{1}{i-1} \sum_{s=k}^{i-1} \left(\frac{s}{i}\right)^a \log\left(\frac{i}{s}\right) \leq 1\,.$$

Integrating the derivative from a to b gives

$$0 \leq p_{n,b}(k) - p_{n,a}(k) \leq b - a \qquad \text{for } 0 \leq a \leq b.$$

To pass from a fixed threshold to the optimum, choose an optimal threshold $k_n(b)$ for parameter b. Then

$$\begin{aligned} \mathrm{OPT}_n(b) &= p_{n,b}(k_n(b)) \\ &\leq p_{n,a}(k_n(b)) + (b - a) \\ &\leq \mathrm{OPT}_n(a) + (b - a). \end{aligned}$$

Also, $\mathrm{OPT}_n(b) \geq \mathrm{OPT}_n(a)$ because every fixed threshold has nondecreasing success probability. Applying these bounds with $a = \alpha_{\mathrm{L}}$ and $b = \alpha$, and using the guarantee already proved, gives

$$0 \leq \mathrm{OPT}_n(\alpha) - P_{\mathcal{D},n}(\alpha) \leq \mathrm{OPT}_n(\alpha) - \mathrm{OPT}_n(\alpha_{\mathrm{L}}) \leq \alpha - \alpha_{\mathrm{L}}.$$

The first inequality holds because no policy can exceed the optimal success probability. Substituting Equation (18), and observing that a difference of success probabilities is at most 1, proves Equation (20). Finally, if $\alpha_{\mathrm{L}} \geq 1$, then $\alpha \geq 1$ and the learned policy is $\mathcal{A}(S)$, which is optimal by Theorem 2. □

Sample size and convergence. For any $\eta \in (0, 1)$, the sufficient condition

$$m \geq \frac{2 \cdot 9^{\alpha}}{\eta^2 (\log 2)^2} \log(2/\delta)$$

makes the loss in Equation (20) at most η, with probability at least $1-\delta$. This follows by substituting the definition of ε_m and solving the inequality $\frac{2 \cdot 3^{\alpha}}{\log 2} \varepsilon_m \leq \eta$ for m. The bound depends on the number of informative training episodes, but not on the test size. For $0 < \alpha \leq 1$, one may replace 9^{α} by 9. One can also retain the observed ratios in the derivative bound: since $t_r^a \geq t_r^{\alpha}$ for $a \leq \alpha$, the same proof allows $3^{-\alpha} \log 2$ to be replaced by $\frac{1}{m} \sum_r t_r^{\alpha} \log\left(\frac{1}{t_r}\right)$.

Finally, consider a growing sequence of independent training episodes with infinitely many episodes satisfying $I_r \geq 2$. Let $\mathcal{D}_m$ contain the first m such episodes, and compute the estimate using

$\delta_m = (m+1)^{-2}$. The theorem holds conditional on every vector of training positions, so its failure probability is also at most δ_m when those positions are random. Thus, for any $M \geq 1$,

$$\mathbb{P}[\text{the theorem's bounds fail for some } m \geq M] \leq \sum_{m=M}^{\infty} \frac{1}{(m+1)^2} \leq \frac{1}{M}.$$

The last inequality follows by comparison with $\int_M^\infty x^{-2}\,dx$. Letting $M \to \infty$ shows that, with probability 1, the bounds hold for all sufficiently large m. For this choice of δ_m, we also have $\varepsilon_m \to 0$. Consequently, $\alpha_{\mathrm{L}} \to \alpha$ and

$$\sup_{n \geq 1} \big(\mathrm{OPT}_n(\alpha) - P_{\mathcal{D}_m,n}(\alpha)\big) \longrightarrow 0 \qquad \text{almost surely.}$$

In other words, with probability 1, the loss tends to zero uniformly over all test sizes. If $\alpha > 1$, convergence of the estimate also gives $\alpha_{\mathrm{L}} > 1$ for all sufficiently large m, so the learned policy is then exactly optimal for every test size.

This construction supplies a conservative estimate; it does not make the unconstrained MLE used in Section F.2 conservative. The guarantees require a common power parameter across training and testing and the stated conditional independence for the test signal. They do not cover family misspecification, rank dependence, or signal corruption.

D Omitted proofs from Section 4

We use the following family in the adversarial-order and multiple-signal proofs. For each $i \in [n]$, let $\mathcal{I}_i := (1, 2, \ldots, i, 0, \ldots, 0) \in \mathbb{N}_0^n$. Its unique maximum arrives at time i, and its first i values agree with those of $\mathcal{I}_n$.

D.1 Randomized optimal policy

Theorem 4. *Let $\alpha > 0$. For adversarial arrival orders with an α-power signal, the optimal worst-case success probability over randomized policies is $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha) = \frac{n^\alpha}{\sum_{j=1}^n j^\alpha}$. The policy $\mathcal{A}(\max\{R, S\})$ is optimal, where $R \in [n]$ is independent of S and, for $r \in [n]$,*

$$\mathbb{P}[R \leq r] = \frac{n^\alpha}{\sum_{j=1}^n j^\alpha} \cdot \frac{\sum_{j=1}^r j^\alpha}{r^\alpha}.$$

Proof. Let

$$c_{n,\alpha} := \frac{n^\alpha}{\sum_{j=1}^n j^\alpha}.$$

We want to show that $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha) = c_{n,\alpha}$. To this end, we first show that $\mathcal{A}(\max\{S, R\})$ guarantees success probability $c_{n,\alpha}$ on every instance of length n before proving that no policy can do better.

We have

$$F(r) := \mathbb{P}[R \leq r] = c_{n,\alpha} \frac{\sum_{j=1}^r j^\alpha}{r^\alpha}, \qquad r \in [n],$$

and set $F(0) := 0$. We first verify that this defines a distribution. Let $H_r := r^{-\alpha} \sum_{j=1}^r j^\alpha$. For every $r \geq 2$ and $j \leq r-1$, we have $\frac{jr}{r-1} \leq j+1$, and hence

$$\begin{aligned} r^\alpha (H_r - H_{r-1}) &= r^\alpha - \sum_{j=1}^{r-1} \left(\left(\frac{jr}{r-1} \right)^\alpha - j^\alpha \right) \\ &\geq r^\alpha - \sum_{j=1}^{r-1} \big((j+1)^\alpha - j^\alpha \big) = 1 > 0. \end{aligned}$$

Thus, F is increasing, and $F(n) = 1$ by the definition of $c_{n,\alpha}$. Therefore, F is a valid cumulative distribution function on $[n]$.

Fix an instance whose maximum arrives at time $i^\star$. Note that $i^\star$ is always a record. Since S follows the α-power distribution on $[i^\star]$, it holds that

$$\mathbb{P}[S = i^\star] = 1 - \left(\frac{i^\star - 1}{i^\star}\right)^\alpha$$

and consequently

$$\mathbb{P}[S \le i^\star - 1] = \left(\frac{i^\star - 1}{i^\star}\right)^\alpha .$$

Policy $\mathcal{A}(\max\{S, R\})$ succeeds in at least the following two disjoint cases: If $S = i^\star$ and $R \le i^\star$ or if $S \le i^\star - 1$ and $R = i^\star$, $\mathcal{A}(\max\{S, R\})$ stops at $i^\star$. Thus,

$$\begin{aligned}
\mathbb{P}[\text{success}] &\ge \left(1 - \left(\frac{i^\star - 1}{i^\star}\right)^\alpha\right) F(i^\star) + \left(\frac{i^\star - 1}{i^\star}\right)^\alpha \big(F(i^\star) - F(i^\star - 1)\big) \\
&= F(i^\star) - \left(\frac{i^\star - 1}{i^\star}\right)^\alpha F(i^\star - 1) \\
&= c_{n,\alpha} \left(\frac{\sum_{j=1}^{i^\star} j^\alpha}{i^{\star\alpha}} - \frac{\sum_{j=1}^{i^\star - 1} j^\alpha}{i^{\star\alpha}}\right) = c_{n,\alpha} .
\end{aligned}$$

Hence, there is a policy with success probability at least $c_{n,\alpha}$, completing the proof of the lower bound on the optimal success probability.

For the matching upper bound, let $\mathcal{A}$ be any randomized policy and let p_i be its success probability on the hard instance $\mathcal{I}_i = (1, 2, \dots, i, 0, \dots, 0)$. With $g_s := s^\alpha - (s-1)^\alpha$ for $s \in [i]$, we have $\mathbb{P}[S = s] = \frac{g_s}{i^\alpha}$ on $\mathcal{I}_i$; for $s > i$, this probability is zero. For a signal time $s \in [n]$ and a time $t \in [n]$, let $q_{s,t}$ be the probability that $\mathcal{A}$ stops at time t on instance $\mathcal{I}_n$ conditioned on the signal occurring at time s. If $s \le i$, then $\mathcal{I}_i$ and $\mathcal{I}_n$ are identical up to time i. Hence, the decisions of $\mathcal{A}$ are identical up to i on $\mathcal{I}_i$ and $\mathcal{I}_n$. Thus, we can rewrite p_i as

$$p_i = \sum_{s=1}^{i} \mathbb{P}\left[\text{success on } \mathcal{I}_i \mid S = s\right] \mathbb{P}\left[S = s\right] = \frac{1}{i^\alpha} \sum_{s=1}^{i} g_s \, q_{s,i}.$$

Therefore,

$$\sum_{i=1}^{n} i^\alpha p_i = \sum_{i=1}^{n} \sum_{s=1}^{i} g_s \, q_{s,i} = \sum_{s=1}^{n} g_s \sum_{i=s}^{n} q_{s,i} \le \sum_{s=1}^{n} g_s = n^\alpha,$$

where the inequality holds because, for each fixed signal time s, the policy $\mathcal{A}$ can stop at most once. This implies

$$\min_{i \in [n]} p_i \le \frac{n^\alpha}{\sum_{i=1}^{n} i^\alpha} = c_{n,\alpha}.$$

Since the worst-case guarantee of $\mathcal{A}$ is at most its minimum success probability on the family $\{\mathcal{I}_i\}_{i=1}^n$, no randomized policy can beat $c_{n,\alpha}$. This completes the proof of the upper bound and thus of the theorem. □

Corollary 5. *Let $(\alpha_n)_{n \ge 1}$ be a positive sequence. If $\alpha_n = o(n)$, then $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) = o(1)$. If $\alpha_n / n \to c \in (0, \infty)$, then $\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) \to 1 - e^{-c}$.*

Proof. The exact formula from Theorem 4 and the bound

$$\sum_{i=1}^{n} i^{\alpha_n} \ge \int_0^n x^{\alpha_n} \, dx = \frac{n^{\alpha_n + 1}}{\alpha_n + 1}$$

give

$$\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) \le \frac{\alpha_n + 1}{n},$$

which proves the first claim when $\alpha_n = o(n)$.

For $\frac{\alpha_n}{n} \to c$, write $\alpha_n = cn + o(n)$ for some constant $c > 0$. Then

$$\sum_{i=1}^{n} \frac{i^{\alpha_n}}{n^{\alpha_n}} = \sum_{i=1}^{n} \left(\frac{i}{n}\right)^{\alpha_n} = \sum_{j=0}^{n-1} \left(1 - \frac{j}{n}\right)^{\alpha_n} .$$

For each fixed j, it holds that

$$\left(1 - \frac{j}{n}\right)^{\alpha_n} = \exp\left(\alpha_n \log\left(1 - \frac{j}{n}\right)\right) = \exp\left(-(c + o(1))j + O\left(\frac{j^2}{n}\right)\right) \xrightarrow{n\to\infty} \exp(-jc) .$$

To justify passage to the infinite sum, note that, for all sufficiently large n, $\frac{\alpha_n}{n} \geq \frac{c}{2}$ and

$$\left(1 - \frac{j}{n}\right)^{\alpha_n} \leq \exp\left(-\frac{\alpha_n j}{n}\right) \leq e^{-cj/2} \qquad (0 \leq j < n).$$

Extending the summand by zero for $j \geq n$, the summable sequence $(e^{-cj/2})_{j\geq 0}$ is a uniform dominating bound. Hence, dominated convergence gives

$$\sum_{i=1}^{n} \frac{i^{\alpha_n}}{n^{\alpha_n}} \xrightarrow{n\to\infty} \sum_{j=0}^{\infty} e^{-cj} = \frac{1}{1 - e^{-c}} .$$

Thus,

$$\mathrm{OPT}_n^{\mathrm{rand}}(\alpha_n) = \frac{n^{\alpha_n}}{\sum_{i=1}^{n} i^{\alpha_n}} = \frac{1}{\sum_{i=1}^{n} i^{\alpha_n}/n^{\alpha_n}} \xrightarrow{n\to\infty} 1 - e^{-c}. \qquad \square$$

D.2 Deterministic optimal policy

Theorem 6. *Let $\alpha > 0$. For adversarial orders with an α-power signal, the policy $\mathcal{A}(S)$ is optimal among deterministic policies, and its worst-case success probability is* $\mathrm{OPT}_n^{\mathrm{det}}(\alpha) = 1 - \left(1 - \frac{1}{n}\right)^{\alpha}$.

Proof. For the lower bound, consider policy $\mathcal{A}(S)$. If the maximum is at position $i^\star$, this policy succeeds if $S = i^\star$, which happens with probability

$$\mathbb{P}[S = i^\star] = 1 - \left(\frac{i^\star - 1}{i^\star}\right)^{\alpha} \geq 1 - \left(1 - \frac{1}{n}\right)^{\alpha} .$$

Hence,

$$\mathrm{OPT}_n^{\mathrm{det}}(\alpha) \geq \min_{i^\star \in [n]} \mathbb{P}[S = i^\star] = 1 - \left(1 - \frac{1}{n}\right)^{\alpha} ,$$

which proves the lower bound on the success probability.

For the upper bound, we have observed that it suffices to consider policies that do not stop before the signal. We may also assume that a policy stops at time n if it has not stopped earlier. Fix such a policy $\mathcal{A}$ and let $\tau : [n] \to [n]$ be its stopping rule on instance $\mathcal{I}_n = (1, 2, \ldots, n-1, n)$. That is, $\tau(s) \geq s$ is the time at which $\mathcal{A}$ stops when the signal occurs at time s. As before, $\mathcal{I}_i$ and $\mathcal{I}_n$ are indistinguishable through time i. On instance $\mathcal{I}_i = (1, 2, \ldots, i, 0, \ldots, 0)$, the success probability is

$$p_i = \mathbb{P}\left[\text{success on } \mathcal{I}_i\right] = \sum_{s=1}^{i} \mathbb{P}\left[\text{success on } \mathcal{I}_i \mid S = s\right] \mathbb{P}\left[S = s\right] = \sum_{s \leq i:\, \tau(s) = i} \frac{s^\alpha - (s-1)^\alpha}{i^\alpha} .$$

If the worst-case guarantee is zero, the desired upper bound is immediate. Suppose, therefore, that it is positive. Then $p_i > 0$ for every $i \in [n]$, so for every i there must exist at least one signal time $s \leq i$ with $\tau(s) = i$. Since there are exactly n signal times and exactly n target times, this implies that every $i \in [n]$ has exactly one preimage under τ. Hence, τ is a bijection of $[n]$. Since $\tau(s) \geq s$ for all s,

$$\sum_{s=1}^{n} \tau(s) \geq \sum_{s=1}^{n} s.$$

Because τ is a bijection, equality must hold. Thus, $\tau(s) = s$ for every $s \in [n]$. Therefore, every deterministic policy with positive worst-case guarantee agrees with $\mathcal{A}(S)$ on the instance family

$\{\mathcal{I}_i\}_{i=1}^n$. In particular, $p_i = \mathbb{P}[\mathcal{A}(S) \text{ succeeds on } \mathcal{I}_i]$. Since $\mathcal{A}(S)$ succeeds on $\mathcal{I}_i$ if and only if $S = i$, this implies

$$p_i = \frac{i^\alpha - (i-1)^\alpha}{i^\alpha} = 1 - \left(\frac{i-1}{i}\right)^\alpha .$$

This quantity is decreasing in i. Thus, the worst-case guarantee is attained at $i = n$ and equals

$$1 - \left(1 - \frac{1}{n}\right)^\alpha .$$

This proves that $\mathcal{A}(S)$ is optimal. □

Corollary 7. *Let $(\alpha_n)_{n\geq 1}$ be a positive sequence. If $\alpha_n = o(n)$, then $\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) = o(1)$. If $\alpha_n/n \to c \in (0,\infty)$, then $\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) \to 1 - e^{-c}$.*

Proof. The exact formula from Theorem 6 gives

$$\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) = 1 - \left(1 - \frac{1}{n}\right)^{\alpha_n} .$$

If $\alpha_n = o(n)$, then $\frac{\alpha_n}{n} \to 0$. Using $\log\left(1 - \frac{1}{n}\right) = -\frac{1}{n} + O\left(\frac{1}{n^2}\right)$, we get

$$\alpha_n \log\left(1 - \frac{1}{n}\right) = -\frac{\alpha_n}{n} + O\left(\frac{\alpha_n}{n^2}\right) = o(1) ,$$

and therefore

$$\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) = 1 - \exp\left(\alpha_n \log\left(1 - \frac{1}{n}\right)\right) = o(1) .$$

If instead $\frac{\alpha_n}{n} \to c \in (0,\infty)$, then

$$\alpha_n \log\left(1 - \frac{1}{n}\right) = \frac{\alpha_n}{n} \cdot n \log\left(1 - \frac{1}{n}\right) \xrightarrow{n\to\infty} -c ,$$

since $n \log\left(1 - \frac{1}{n}\right) \xrightarrow{n\to\infty} -1$. Hence,

$$\mathrm{OPT}_n^{\mathrm{det}}(\alpha_n) = 1 - \exp\left(\alpha_n \log\left(1 - \frac{1}{n}\right)\right) \xrightarrow{n\to\infty} 1 - e^{-c} .$$ □

E Omitted proofs from Section 5

A full signal history is a vector

$$h = (c_1, \ldots, c_{\ell(h)}, 0, \ldots, 0) \in \{0, \ldots, m\}^n,$$

where c_t is the number of signals at time t, $c_{\ell(h)} > 0$, and $\sum_{t=1}^{\ell(h)} c_t = m$. Thus, $\ell(h)$ is the time of the last signal. Let $\mathcal{H}_m$ be the set of all such histories; below, we suppress trailing zeros when writing its elements. Setting $0! := 1$, define

$$\lambda_i(h) = \begin{cases} \dfrac{m!}{c_1! \cdots c_{\ell(h)}!} \dfrac{1}{i^m}, & \ell(h) \leq i, \\ 0, & \ell(h) > i. \end{cases}$$

Lemma 17. *Each history $h \in \mathcal{H}_m$ is observed on $\mathcal{I}_i$ with probability $\lambda_i(h)$.*

Proof. Fix an instance $\mathcal{I}_i$. In the full-history model, the m signals are drawn independently and uniformly from $[i]$. It is convenient to temporarily consider the signals as labeled, so a signal realization is a tuple $s = (s_1, \ldots, s_m) \in [i]^m$. Each such labeled tuple has probability i^{-m}.

Given a labeled tuple s, the observed history records only the number of signals at each time. That is, if $\ell(s) := \max_{j\in[m]} s_j$, then the induced history is $h(s) = (c_1(s), \ldots, c_{\ell(s)}(s))$, where $c_t(s) =$

$|\{j \in [m] : s_j = t\}|$. Thus the policy observes the counts of signals at each time, but not the labels of the signals. Now fix a history $h = (c_1, \dots, c_{\ell(h)}) \in \mathcal{H}_m$. If $\ell(h) > i$, then no signal can occur at time $\ell(h)$ on the instance $\mathcal{I}_i$ since all signals are supported on $[i]$. Therefore h is impossible, and

$$\mathbb{P}[h \text{ is observed on } \mathcal{I}_i] = 0 = \lambda_i(h).$$

Assume instead that $\ell(h) \le i$. The history h is observed exactly when, for every $t \in [\ell(h)]$, precisely c_t of the labeled signals are equal to t. Since $\sum_{t=1}^{\ell(h)} c_t = m$, this also accounts for all m signals, so no signal occurs after time $\ell(h)$.

It remains to count how many labeled tuples $s \in [i]^m$ induce this same unlabeled history. We first choose the c_1 labeled signals that occur at time 1, then the c_2 labeled signals among the remaining ones that occur at time 2, and so on. This gives

$$\binom{m}{c_1}\binom{m-c_1}{c_2}\binom{m-c_1-c_2}{c_3}\cdots\binom{c_{\ell(h)}}{c_{\ell(h)}}.$$

The factorials telescope, so the number of labeled tuples producing h is

$$\frac{m!}{c_1!\cdots c_{\ell(h)}!}.$$

Here the convention $0! = 1$ handles times at which no signal occurs.

Since each labeled tuple has probability i^{-m}, we obtain

$$\mathbb{P}[h \text{ is observed on } \mathcal{I}_i] = \frac{m!}{c_1!\cdots c_{\ell(h)}!} \cdot \frac{1}{i^m}.$$

This is exactly $\lambda_i(h)$ by definition. Hence every history $h \in \mathcal{H}_m$ is observed on $\mathcal{I}_i$ with probability $\lambda_i(h)$. □

E.1 Randomized optimal policy

Theorem 8. *For all integers $n, m \ge 1$, it holds that* $\mathrm{OPT}_n^{\mathrm{rand}}(m) = \mathrm{OPT}_n^{\mathrm{rand,full}}(m)$.

Proof. The inequality $\mathrm{OPT}_n^{\mathrm{rand}}(m) \le \mathrm{OPT}_n^{\mathrm{rand,full}}(m)$ is immediate, since a full-history policy can ignore all signals except the last one.

For the reverse inequality, fix a full-history policy and any arrival order with best-item position i. Lemma 17 gives the probability for observing a particular history $h = (c_1, \dots, c_\ell)$ with last signal at ℓ. Since $\mathbb{P}[L = \ell] = \frac{\ell^m - (\ell-1)^m}{i^m}$, we obtain Equation (6):

$$\mathbb{P}[h \mid L = \ell, i^\star = i] = \frac{m! / \prod_{t=1}^{\ell} c_t!}{\ell^m - (\ell-1)^m} \tag{6}$$

This is independent of i and hence of the fixed arrival order.

Using internal randomness, we can now simulate a policy with access to the full signal history: A randomized policy observing only L rejects until L arrives. At that time it samples a history from Equation (6) and replays the full-history policy on the stored relative-rank observations, with additional, but independent internal randomness. If the simulated policy selected an item before L, the new policy makes no selection; that earlier selection could not have succeeded. Otherwise, it follows the simulated policy from time L onward, including its decision at L. All signals have then arrived, so no future signal observations need to be generated.

Conditional on L, the simulated history has the same distribution as the actual full history on every fixed arrival order. The two policies therefore have identical success probabilities on every such order. This proves $\mathrm{OPT}_n^{\mathrm{rand,full}}(m) \le \mathrm{OPT}_n^{\mathrm{rand}}(m)$. □

E.2 Deterministic optimal policy

For $h \in \mathcal{H}_m$ and $t \in \{\ell(h), \dots, n\}$, the variable $x_{h,t} = 1$ means that the policy stops at time t after observing history h; the objective z is the guaranteed success probability. Recall the integer linear program from Section 5:

$$\begin{aligned} \max \quad & z && && \text{(ILP)} \\ \text{s.t.} \quad & \sum_{t=\ell(h)}^{n} x_{h,t} = 1 && \forall h \in \mathcal{H}_m \\ & \sum_{h \in \mathcal{H}_m : \ell(h) \le i} \lambda_i(h)\, x_{h,i} \ge z && \forall i \in [n] \\ & x_{h,t} \in \{0,1\} && \forall h \in \mathcal{H}_m,\ \forall t \in \{\ell(h), \dots, n\}. \end{aligned}$$

Theorem 9. *For all $m, n \in \mathbb{N}$, the optimal objective value of* (ILP) *is* $\mathrm{OPT}_n^{\mathrm{det,full}}(m)$.

Proof. First consider any deterministic policy $\mathcal{A}$. As observed before, we may assume that $\mathcal{A}$ stops only after all m signals have appeared and stops at time n if it has not stopped earlier. Consider its behavior on $\mathcal{I}_n$. For every history h, let $\tau(h) \in \{\ell(h), \dots, n\}$ be the stopping time of $\mathcal{A}$ after observing h, and set $x_{h,t} = \mathbb{1}[\tau(h) = t]$. Then the first set of constraints is satisfied.

Now fix $i \in [n]$. Up to time i, the two instances $\mathcal{I}_i$ and $\mathcal{I}_n$ are identical, so after any history h with $\ell(h) \le i$, the policy makes the same decision on $\mathcal{I}_i$ as on $\mathcal{I}_n$. Hence, by Lemma 17 the success probability of $\mathcal{A}$ on $\mathcal{I}_i$ is exactly

$$\sum_{h : \ell(h) \le i} \lambda_i(h)\, x_{h,i}.$$

Therefore every deterministic policy induces a feasible solution to (ILP) with

$$z \le \min_{i \in [n]} \mathbb{P}[\mathcal{A} \text{ wins on } \mathcal{I}_i].$$

Since the worst-case guarantee of $\mathcal{A}$ is at most its minimum success probability on the hard family $\{\mathcal{I}_i\}_{i=1}^n$, we obtain that $\mathrm{OPT}_n^{\mathrm{det,full}}(m)$ is at most the optimal objective value of (ILP).

Conversely, let x be any feasible solution to (ILP) with value z. Define a deterministic policy that ignores the item values, waits until all m signals have appeared, reads the resulting history h, and then stops at the unique time t with $x_{h,t} = 1$. For any adversarial instance whose maximum is at position i, the distribution of the signal history is the same as on $\mathcal{I}_i$ because the signals depend only on the maximum position. Thus, the success probability of this policy is exactly

$$\sum_{h :\, \ell(h) \le i} \lambda_i(h) \cdot x_{h,i}$$

when the maximum is at position i, and this is at least z for every $i \in [n]$ by the second set of constraints. Hence the policy has worst-case guarantee at least z. This shows that the ILP optimum is at most $\mathrm{OPT}_n^{\mathrm{det,full}}(m)$. □

Using the characterization just proved, we now turn to the special case of two signals.

Theorem 18. *For $m = 2$, we have*

$$\mathrm{OPT}_n^{\mathrm{det,full}}(2) = \max\Big\{ z \in [0,1] : \sum_{t=1}^{\ell} \lceil z t^2 \rceil \le \ell^2 \text{ for all } \ell \in [n] \Big\}.$$

Proof. For $m = 2$, write a history as a sorted pair (a, b), $1 \le a \le b \le n$, and let

$$w(a,b) = \begin{cases} 2, & a < b, \\ 1, & a = b. \end{cases}$$

The weight $w(a,b)$ is the number of *labeled* signal pairs inducing the sorted history (a,b). We may assume that a deterministic policy waits until both signals have appeared and stops at time n if it has not stopped earlier.

In order to show

$$\mathrm{OPT}_n^{\mathrm{det,full}}(2) \le \max\Big\{z \in [0,1] : \sum_{t=1}^{\ell} \lceil zt^2 \rceil \le \ell^2 \text{ for all } \ell \in [n]\Big\},$$

fix a deterministic policy and consider its behavior on the hard instance $\mathcal{I}_n$. For every history (a,b), let $\tau(a,b) \ge b$ be the stopping time chosen by the policy on $\mathcal{I}_n$. Define

$$C_t := \sum_{\substack{1 \le a \le b \le t \\ \tau(a,b)=t}} w(a,b).$$

On the hard instance $\mathcal{I}_t$, the history (a,b) occurs with probability $\frac{w(a,b)}{t^2}$, and the policy wins exactly when $\tau(a,b) = t$. Hence its success probability on $\mathcal{I}_t$ is $\frac{C_t}{t^2}$.

Suppose the policy has worst-case guarantee at least z. Then $C_t \ge zt^2$, and since C_t is an integer, $C_t \ge \lceil zt^2 \rceil$. Moreover, for every ℓ, all histories relevant for $\mathcal{I}_\ell$ must have second signal at most ℓ. The total weight of such histories is

$$\sum_{b=1}^{\ell} \sum_{a=1}^{b} w(a,b) = \sum_{b=1}^{\ell} (2b-1) = \ell^2.$$

Therefore

$$\sum_{t=1}^{\ell} \lceil zt^2 \rceil \le \sum_{t=1}^{\ell} C_t \le \ell^2$$

for all $\ell \in [n]$. This proves the upper bound.

For the reverse inequality, suppose that $z \in [0,1]$ satisfies

$$\sum_{t=1}^{\ell} \lceil zt^2 \rceil \le \ell^2$$

for all $\ell \in [n]$. We construct a deterministic policy with worst-case success probability at least z. Set $d_t := \lceil zt^2 \rceil$ for all $t < n$ and $d_n := n^2 - \sum_{t=1}^{n-1} d_t$. The constraint for $\ell = n$ implies $d_n \ge \lceil zn^2 \rceil$.

We assign histories greedily to stopping times so that the total weight assigned to time t is d_t. Suppose that the assignments for times $1, \dots, t-1$ have already been made, and let

$$\mathcal{A}_t := \{(a,b) : 1 \le a \le b \le t \text{ and } (a,b) \text{ is not yet assigned}\}$$

be the set of currently available histories. Every history in $\mathcal{A}_t$ can be assigned to time t, since its second signal is at most t. Moreover, the total available weight is

$$W_t := \sum_{(a,b) \in \mathcal{A}_t} w(a,b) = t^2 - \sum_{s=1}^{t-1} d_s \ge d_t,$$

where the inequality follows from the prefix constraint.

It remains to justify that one can select available histories of *exactly* weight d_t. The available histories have weights only 1 and 2, and the diagonal history (t,t) is newly available at step t and was not available before, so it is still unassigned and has weight 1. Thus $\mathcal{A}_t$ contains at least one weight-1 history. Let $p \ge 1$ be the number of weight-1 histories in $\mathcal{A}_t$, and let q be the number of weight-2 histories. Then $W_t = p + 2q$. We claim that for every integer $r \in \{0, \dots, W_t\}$ there is a subset $B \subseteq \mathcal{A}_t$ with

$$\sum_{(a,b) \in B} w(a,b) = r.$$

Indeed, if $r \le 2q$, write $r = 2u + \varepsilon$ with $\varepsilon \in \{0, 1\}$. If $\varepsilon = 0$, choose $u \le q$ weight-2 histories. If $\varepsilon = 1$, then $r \le 2q$ implies $u \le q - 1$, so choose u weight-2 histories and one weight-1 history. If instead $r > 2q$, choose all q weight-2 histories and $r - 2q$ weight-1 histories; this is possible because $r \le p + 2q$ implies $r - 2q \le p$.

Applying the claim with $r = d_t$ gives a subset of available histories of total weight exactly d_t, which we assign to stopping time t. We do this for every $t < n$. At time n, assign all remaining histories to time n. The total remaining weight is

$$n^2 - \sum_{t=1}^{n-1} d_t = d_n,$$

so the weight assigned to time n is exactly d_n. Let $\tau(a, b)$ denote the stopping time assigned to history (a, b). By construction, for every $t \in [n]$, the total weight of histories assigned to stopping time t is exactly d_t, that is,

$$\sum_{\substack{1 \le a \le b \le t \\ \tau(a,b) = t}} w(a, b) = d_t.$$

Now consider any instance whose maximum arrives at time t. The two signals are independent and uniform on $[t]$. Hence a sorted history (a, b) with $a \le b \le t$ occurs with probability $\frac{w(a,b)}{t^2}$. The policy wins exactly on those histories that are assigned to stopping time t because then it stops exactly when the maximum arrives. Therefore the success probability on such an instance is

$$\sum_{\substack{1 \le a \le b \le t \\ \tau(a,b) = t}} \frac{w(a, b)}{t^2} = \frac{d_t}{t^2} \ge z \ ,$$

where the final inequality follows from $d_t = \lceil zt^2 \rceil \ge zt^2$ for $t < n$, and

$$d_n = n^2 - \sum_{t=1}^{n-1} \lceil zt^2 \rceil \ge \lceil zn^2 \rceil \ge zn^2 \ .$$

This defines a deterministic policy with worst-case success probability at least z. Since z is achievable by a deterministic policy, $\mathrm{OPT}_n^{\mathrm{det,full}}(2)$ must be at least z. This concludes the proof. □

Corollary 19. *For every integer $n \ge 4$, we have $\frac{6(n-1)}{(n+1)(2n+1)} \le \mathrm{OPT}_n^{\mathrm{det,full}}(2) \le \frac{6n}{(n+1)(2n+1)}$. In particular, it holds that $\mathrm{OPT}_n^{\mathrm{det,full}}(2) = \frac{3}{n} + O\left(\frac{1}{n^2}\right)$.*

Proof. Let $z^\star := \mathrm{OPT}_n^{\mathrm{det,full}}(2)$. Then, $z^\star$ is feasible for the characterization given in Theorem 18. Hence, using the constraint for $\ell = n$, we get

$$z^\star \sum_{t=1}^{n} t^2 \le \sum_{t=1}^{n} \lceil z^\star t^2 \rceil \le n^2.$$

Since $\sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6}$, this gives

$$\mathrm{OPT}_n^{\mathrm{det,full}}(2) = z^\star \le \frac{n^2}{\sum_{t=1}^{n} t^2} = \frac{6n}{(n+1)(2n+1)}.$$

For the lower bound, set

$$\underline{z} := \frac{6(n-1)}{(n+1)(2n+1)}.$$

We show that $\underline{z}$ is feasible for the characterization given in Theorem 18. For this, set

$$g(\ell) := \frac{6(\ell-1)}{(\ell+1)(2\ell+1)}.$$

A direct computation gives, for every $\ell \geq 3$,

$$g(\ell+1) - g(\ell) = \frac{12(-\ell^2+\ell+3)}{(\ell+1)(\ell+2)(2\ell+1)(2\ell+3)} < 0.$$

Thus g is decreasing on $\{3,4,\ldots\}$. Moreover, $g(2) = g(4) = \frac{2}{5}$, so for every $n \geq 4$ and every $\ell \in \{2,\ldots,n\}$ we have

$$\underline{z} = g(n) \leq g(\ell).$$

Now fix $\ell \in [n]$. If $\ell = 1$, then $0 < \underline{z} \leq 1$, and hence

$$\sum_{t=1}^{1} \lceil \underline{z} t^2 \rceil = \lceil \underline{z} \rceil = 1.$$

If $\ell \geq 2$, then $\lceil x \rceil \leq x+1$ for every real x, and therefore

$$\sum_{t=1}^{\ell} \lceil \underline{z} t^2 \rceil \leq \underline{z} \sum_{t=1}^{\ell} t^2 + \ell = \underline{z} \frac{\ell(\ell+1)(2\ell+1)}{6} + \ell \leq g(\ell) \frac{\ell(\ell+1)(2\ell+1)}{6} + \ell = \ell^2.$$

Thus $\underline{z}$ satisfies all feasibility inequalities in Theorem 18, and so

$$\mathrm{OPT}_n^{\mathrm{det,full}}(2) \geq \underline{z} = \frac{6(n-1)}{(n+1)(2n+1)}.$$

The two bounds differ by

$$\frac{6n}{(n+1)(2n+1)} - \frac{6(n-1)}{(n+1)(2n+1)} = O\left(\frac{1}{n^2}\right),$$

and the upper bound is $\frac{3}{n} + O\big(\frac{1}{n^2}\big)$. Hence,

$$\mathrm{OPT}_n^{\mathrm{det,full}}(2) = \frac{3}{n} + O\left(\frac{1}{n^2}\right). \qquad \square$$

F Additional numerical experiments

F.1 Additional random-order experiments

This subsection gives the protocol for the left and right panels of Figure 3 and the additional random-order comparisons. In each trial we sample a full random permutation of ranks $[n]$, draw the signal S conditional on the realized arrival time I of the maximum, run the corresponding stopping rule on the realized order, and record whether the selected item is the maximum. Each empirical success probability averages 1000 trials; confidence bands are the mean ± 1.96 standard errors. The middle panel of Figure 3 instead evaluates learned policies by their exact test success probabilities, as described in Section F.3. The separate held-out fitting study is in Section F.2.

Protocol for the main comparisons. In Figure 3, $n = 1000$ and clean signals follow the α-power distribution conditional on I. Let k_n^{cl} be the smallest threshold maximizing the classic success probability, where $p_n^{\mathrm{cl}}(1) = \frac{1}{n}$ and $p_n^{\mathrm{cl}}(k) = \frac{k-1}{n} \sum_{i=k}^{n} \frac{1}{i-1}$ for $k \geq 2$. The figure legends use $\frac{n}{e}$ as shorthand for this exact threshold. The left panel uses 41 logarithmically spaced values of α from 0.05 to 100, with all policies sharing the sampled instances at each value. In the right panel, $\alpha = 1$ and corruption occurs independently with probability ρ, using 41 equally spaced values from 0 to 1. Conditional on corruption, the signal is missed, replaced by a uniformly random false alarm in $[n]$, or delayed to a uniformly random time after I, each with equal probability. When $I = n$, the delayed branch is treated as unobserved. The fallback accepts the first record at or after the observed signal; if no signal has arrived by k_n^{cl}, it uses the classic rule.

Model scaling with n. The first experiment repeats the model evaluation for $n \in \{5, 20, 100\}$ and 13 logarithmically spaced values of α from 0.05 to 10. For each α, we use the finite-n optimal threshold from Theorem 2 and estimate its success probability. Figure 4 compares these empirical curves with the asymptotic optimum $\mathrm{OPT}(\alpha)$ and the classic limit $\frac{1}{e}$. The curves approach the asymptotic formula as n grows. The left panel of Figure 3 separately compares against the exact finite-n classic policy.

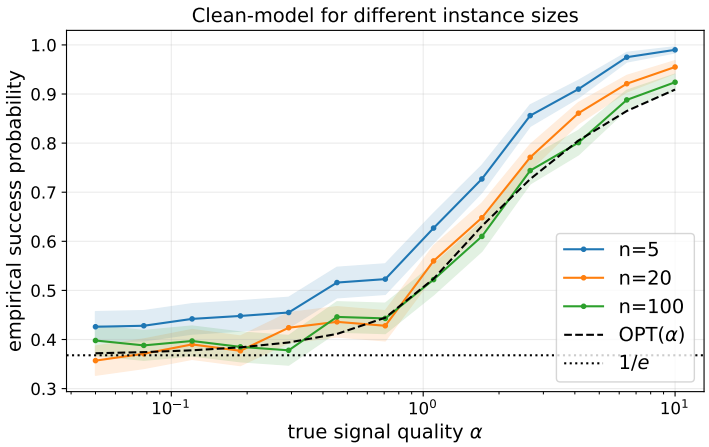

Figure 4: Random-order experiments for several problem sizes. Each curve shows the empirical success probability of the finite-n optimal asynchronous threshold policy from Theorem 2. The dashed curve is the asymptotic optimum $\mathrm{OPT}(\alpha)$.

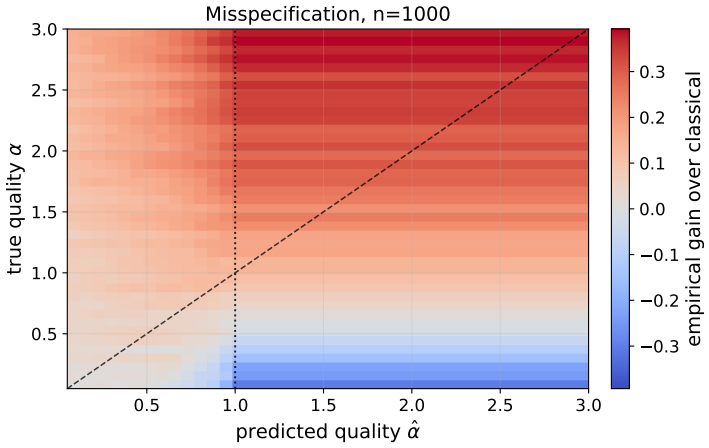

Figure 5: Empirical gain over the finite-n classic baseline under misspecified signal quality at $n = 1000$. The dashed diagonal marks $\hat{\alpha} = \alpha$; all predictions on or to the right of the vertical line $\hat{\alpha} = 1$ induce $\mathcal{A}(S)$.

Misspecification slices and induced thresholds. Figure 5 summarizes robustness to misspecification at $n = 1000$. Both the true parameter α and the predicted parameter $\hat{\alpha}$ range over 41 equally spaced values from 0.05 to 3. The heatmap shows empirical gain over the classic policy; paired standard errors accompany the supplementary data. The dashed diagonal marks $\hat{\alpha} = \alpha$, and the vertical line marks $\hat{\alpha} = 1$. Figure 6 gives a complementary one-dimensional view at the same n, with true parameters $\alpha \in \{0.25, 0.5, 1, 2\}$ and 31 equally spaced predictions from 0.05 to 1.5. We consider the policy $\mathcal{A}(\max\{S, \widehat{k}_n\})$ from Theorem 3 that uses the additional threshold $\widehat{k}_n = \lceil \beta^\star(\hat{\alpha})n \rceil$. For each true α, the same sampled instances are used for all values of $\hat{\alpha}$, giving a paired comparison across thresholds. The lower panel of Figure 6 plots the induced threshold fraction $\frac{\widehat{k}_n}{n}$. This makes explicit why the success curves are flat once $\hat{\alpha} \geq 1$: all such predictions induce the same signal-trusting policy $\mathcal{A}(S)$. The qualitative behavior is consistent with Theorem 3: asymptotically, conservative threshold choice retains the classic benchmark, while sufficiently large overestimates can be harmful when the true signal tends to arrive early.

Separating the noisy-signal failure modes. The right panel of Figure 3 combines three corruption types into one mixed corruption rate. Figure 7 separates these effects, again with $n = 1000$ and

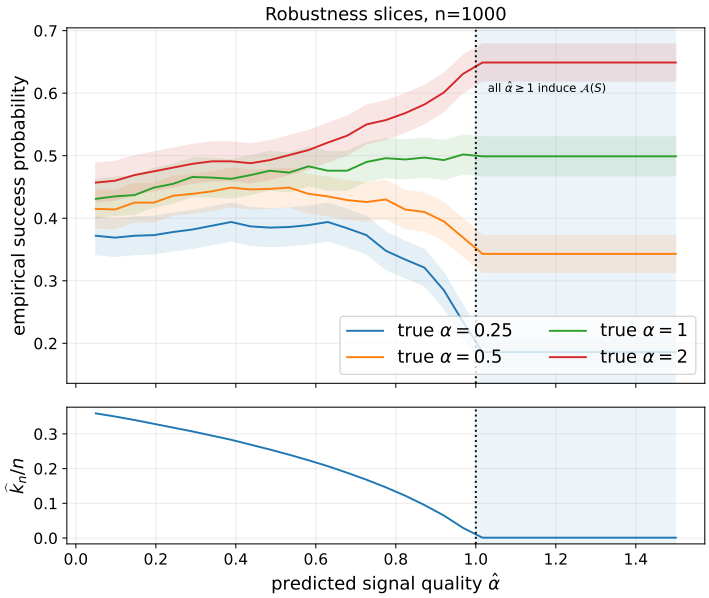

Figure 6: Misspecification slices in the random-order model. The top panel shows empirical success probabilities as a function of the predicted parameter $\hat{\alpha}$ for several true values of α. The bottom panel shows the induced threshold fraction $\widehat{k}_n/n$, explaining the plateau for $\hat{\alpha} \geq 1$.

$\alpha = 1$, using 17 equally spaced values of ρ from 0 to 1. In each panel, with probability $1 - \rho$ the clean signal is observed, while with probability ρ the indicated corruption occurs: the signal is missed, replaced by a uniformly random false alarm in $[n]$, or delayed to a uniformly random time after I. In the delayed branch, $I = n$ is treated as an unobserved signal. We compare the clean-model-optimal policy from Theorem 2, $\mathcal{A}(k_n^{\mathrm{cl}})$, and the fallback policy from Section 6. Missed and late signals mostly harm the clean-model policy by preventing it from acting in time, whereas false alarms can induce premature stopping. The fallback rule mitigates missing and late signals by activating the classic threshold when the signal is absent or arrives too late; early false alarms remain a failure mode.

Exact behavior of the fallback under complete corruption. Write $k = k_n^{\mathrm{cl}}$ and let $p_n^{\mathrm{cl}}(k)$ be the classic success probability defined above. For $\alpha = 1$, the fallback starts accepting records at $\min\{S_{\mathrm{obs}}, k\}$, with an unobserved signal represented by $n + 1$. At $\rho = 1$, the missing and late branches each have classic success probability: a late signal preceding k only occurs on instances with $I < k$, where the classic policy also fails. In the false-alarm branch, the start time is s with probability $\frac{1}{n}$ for each $s < k$, and is k with probability $\frac{n-k+1}{n}$. Thus the mixed success probability is exactly

$$\frac{2}{3}p_n^{\mathrm{cl}}(k) + \frac{1}{3n}\left(\sum_{s=1}^{k-1} p_n^{\mathrm{cl}}(s) + (n-k+1)p_n^{\mathrm{cl}}(k)\right).$$

For $n = 1000$ and $k = 369$, this is 0.35686775, compared with $p_n^{\mathrm{cl}}(k) = 0.36819562$. Since $\frac{k}{n} \to \beta = \frac{1}{e}$ and $p_n^{\mathrm{cl}}(\lceil xn \rceil) \to -x \log x$, the false-alarm branch tends to

$$\int_0^\beta -x \log x \, dx + (1-\beta)(-\beta \log \beta) = \frac{1}{e} - \frac{1}{4e^2}.$$

The equal mixture therefore tends to $\frac{1}{e} - \frac{1}{12e^2}$. For an arbitrary false alarm at time 1, the fallback instead accepts the first item and succeeds with probability only $\frac{1}{n}$.

A provable corruption safeguard. Under the independent-contamination model used in this experiment, choose the classic policy with probability λ and the clean-signal policy otherwise. If

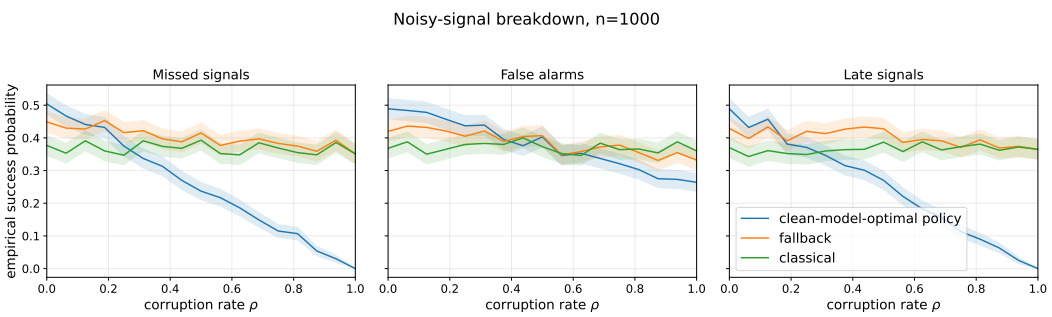

Figure 7: Noisy-signal breakdown in the random-order model. The three panels isolate missed signals, false alarms, and late signals. The fallback policy gives a smoother degradation than the clean-model-optimal asynchronous policy across the different corruption mechanisms.

the signal is drawn from the clean signal distribution with probability $1-\rho$ and only the signal is corrupted otherwise, the success probability is at least

$$\lambda p_{\text{classic}} + (1-\lambda)(1-\rho)p_{\text{signal}},$$

where p_{classic} and p_{signal} are the respective clean-model success probabilities. The guarantee remains at least $\lambda p_{\text{classic}}$ even if the contaminated branch is arbitrary. This is only a baseline, not an optimal policy for noisy signals.

F.2 Learning the signal model and testing its assumptions

Dependence with unchanged signal marginals. The conditional-independence assumption cannot be checked by fitting $q_i(s)$ alone. For $I = i \geq 2$, let J be the position of the best item among the first $i-1$ arrivals. Set $S = J$ with probability $\frac{i-1}{i}$ and $S = i$ otherwise, using an independent coin; for $i = 1$, set $S = 1$. Since J is uniform on $[i-1]$, this construction has $S \mid I = i \sim \mathcal{U}([i])$ and $S \leq I$, exactly as in the uniform model. Nevertheless, when $S = J$, the signal arrives at a record preceding the best item, so $\mathcal{A}(S)$ stops too early. It succeeds only when $S = i$, giving

$$\mathbb{P}[\mathcal{A}(S) \text{ succeeds}] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i} = \frac{H_n}{n} \longrightarrow 0.$$

More generally, mixing this construction with an independent uniform signal with probabilities η and $1-\eta$ preserves every $q_i(s)$, while the success probability becomes $(1-\eta)\frac{(n+1)}{2n} + \eta\frac{H_n}{n}$. Thus, conditional independence is plausible for a controlled signal generator using only I and fresh randomness, but requires separate validation for signals derived from candidate quality or observed activity.

Fitting on historical episodes. We fix $n = 1000$ and repeat each experiment 30 times. Each repetition uses 200 independent training episodes and 2000 fresh test episodes; all policies share the test episodes. We assume historical episodes reveal both I and S after completion. From the training pairs (I_r, S_r), we fit the α-power model by maximizing the exact discrete conditional log likelihood

$$\sum_{r=1}^{200}\log\left(\frac{S_r^a - (S_r-1)^a}{I_r^a}\right), \qquad a > 0,$$

and use the resulting estimate $\hat{\alpha}$ in $\mathcal{A}(\max\{S, \widetilde{k}_n(\beta^\star(\hat{\alpha}))\})$. This study uses the rounded asymptotic threshold, whereas the learning curves in Section F.3 use exact finite-n thresholds for both estimators. Numerical optimization uses $e^{-7} \leq a \leq e^{7}$; no reported fit reaches either endpoint. The estimate is fitted without a conservativeness constraint, so Theorem 3 does not provide a finite-sample guarantee for this procedure. The confidence-bound estimator in Section C.1 supplies such a guarantee for a different procedure, evaluated separately in Section F.3. We compare MLE with $\mathcal{A}(\max\{S, k_n^{\text{cl}}\})$, which requires no estimate, and the classic policy. All episodes use full uniformly random permutations. Besides power signals with $\alpha = 0.3$ and $\alpha = 1$, we test an equal mixture of

Table 2: Held-out success probabilities. The second column is the median fitted parameter; the remaining columns report means with approximate 95% intervals (mean ± 1.96 standard errors across 30 independent repetitions). "No tuning" uses $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$ without estimating α.

Signal model	$\hat{\alpha}$	Fitted policy	No tuning	classic
Power, $\alpha = 0.3$	0.299	0.397 ± 0.004	0.391 ± 0.005	0.369 ± 0.005
Power, $\alpha = 1$	1.006	0.500 ± 0.004	0.428 ± 0.004	0.364 ± 0.004
Equal power mixture	0.456	0.480 ± 0.004	0.445 ± 0.004	0.371 ± 0.004
Dependent uniform	0.990	0.061 ± 0.020	0.369 ± 0.005	0.368 ± 0.005

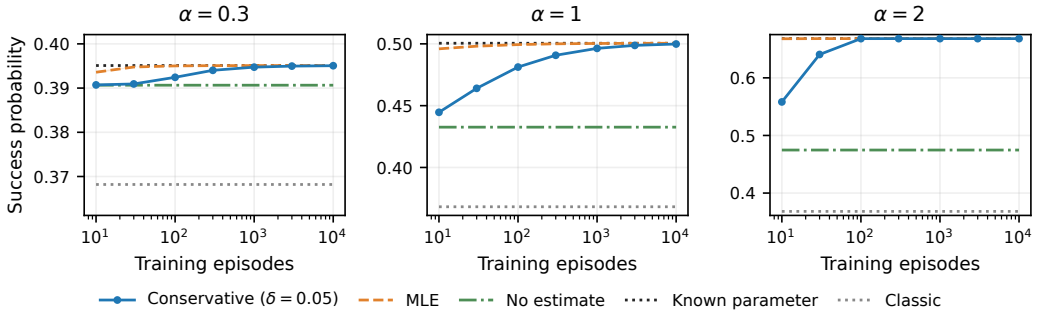

Figure 8: Learning curves at $n = 1000$: exact test success averaged over 1000 independent training histories. Shading gives approximate 95% confidence intervals for the means. The $\alpha = 1$ comparison also appears in the middle panel of Figure 3.

power signals with parameters 0.2 and 3 (outside the fitted family but conditionally independent), and the dependent uniform construction above. The mixture component is drawn independently in each episode; training and test episodes always follow the same specified model.

Table 2 separates the value of fitting from that of the reliable lower bound alone. Fitting improves over the "No tuning" baseline in the independent models tested, including the mixture, although this is not a general guarantee under family misspecification. In the dependent model, $\hat{\alpha}$ remains close to 1 while the fitted policy's success falls to about 0.061. The signal-only rule succeeds with empirical probability 0.00745, consistent with $\frac{H_{1000}}{1000} \approx 0.00749$. The "No tuning" baseline retains the classic success probability, as its realization-wise guarantee uses only $S \le I$. These synthetic tests isolate assumptions and do not establish the suitability of the model for a particular application.

F.3 Learning curves for the conservative estimator

Training and policy choice. We evaluate the estimator in Theorem 16 for $\alpha \in \{0.3, 1, 2\}$, with training and test length $n = 1000$ and failure probability $\delta = 0.05$. For each parameter, we generate 1000 independent training histories, each containing 10,000 episodes. In each episode, we draw I uniformly from $[n]$ and draw S independently of relative ranks conditional on I, using the α-power distribution. These pairs have exactly the same distribution as those obtained from full random permutations; the remaining ranks are not used in fitting.

We use prefixes of length $N \in \{10, 30, 100, 300, 1000, 3000, 10000\}$ from each history. Both estimators receive the same pairs, and neither uses test data. The conservative estimate discards episodes with $I = 1$ and uses the remaining count m in its confidence bound; the horizontal axis in Figure 8 shows the total count N before this filtering. MLE maximizes the exact discrete conditional likelihood as in Section F.2, with the same search interval $e^{-7} \le a \le e^{7}$; no reported fit reaches an endpoint. To compare the estimators directly, both policies use the exact finite-n optimal threshold for their estimates. For $\alpha_{\mathrm{L}} = 0$, we use k_n^{cl}. The *No estimate* baseline uses $\mathcal{A}(\max\{S, k_n^{\mathrm{cl}}\})$, the *Classic* baseline ignores the signal, and the *Known parameter* baseline uses the optimal threshold for the true α.

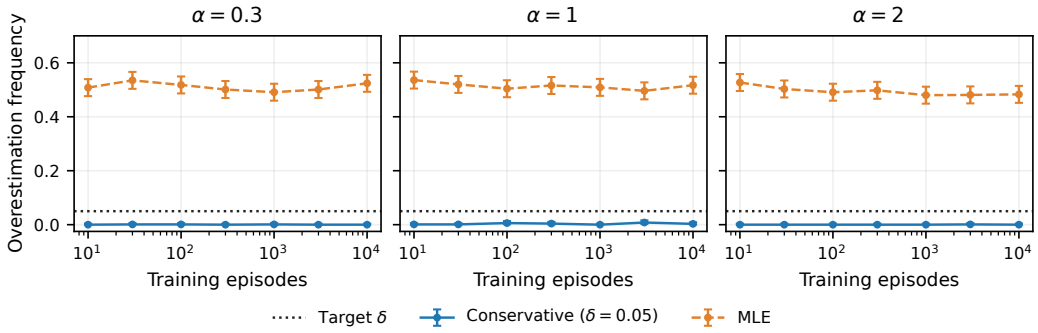

Figure 9: Frequency of estimates exceeding the true parameter across 1000 independent training histories at each size. Error bars are pointwise exact 95% binomial confidence intervals; the dotted line is $\delta = 0.05$. This checks parameter conservativeness, which is distinct from falling below the classic success probability.

Evaluation and uncertainty. For every fitted policy, we compute its success probability on a fresh random-order instance exactly, using the finite sum in the proof of Theorem 16. We then average these probabilities across the 1000 training histories. Thus, variation in Figure 8 comes from the training data, rather than from a finite test sample. Shaded intervals are the mean ± 1.96 standard errors across independent histories. Histories are independent within each training-size comparison; points at different sizes share prefixes.

Performance and conservativeness. The conservative policy improves with more training data in all three cases. For $\alpha = 0.3$, its mean success increases from 0.39072 at $N = 10$ to 0.39506 at $N = 10{,}000$, compared with the optimum 0.39510. For $\alpha = 1$, the corresponding values are 0.44463 and 0.49988, compared with 0.50050. For $\alpha = 2$, all 1000 conservative estimates are at least 1 by $N = 300$ and at every larger size tested, so their policies are exactly optimal. MLE approaches the optimum faster on average. This comparison shows the cost of requiring a conservative estimate, rather than an advantage in average performance over MLE.

Figure 9 reports the fraction of histories in which each estimate exceeds the true α. Conservative overestimation occurs in at most 8 of 1000 histories at any tested parameter and size, below the target probability $\delta = 0.05$. MLE overestimates in approximately half of the histories. Overestimation need not cause a loss below the classic benchmark: for example, every estimate at least 1 induces the optimal policy when $\alpha \geq 1$. Nevertheless, at $\alpha = 0.3$ and $N = 10$, the MLE policy's true success probability is below the classic benchmark in 7 histories. No such violation occurs for the conservative policy in the reported runs. These are empirical frequencies for the specified model; the finite-sample guarantee comes from Theorem 16.

F.4 Additional adversarial-order experiments

We now supplement the adversarial-order results from Section 4. These experiments use the hard family $\mathcal{I}_i = (1, 2, \ldots, i, 0, \ldots, 0)$ where the maximum is at position i and no item after time i is a record. Unlike the random-order experiments above, there is no permutation randomness: the adversary fixes the instance $\mathcal{I}_i$. In the single-signal experiments, randomness comes only from the signal time S and, for randomized policies, from the internal random threshold of the algorithm. The full-history experiment for $m = 2$ is computed numerically from the exact characterization in Theorem 18.

Position profile of the hard family. The first adversarial experiment fixes $n = 1000$ and $\alpha = 1000$ and varies the adversarial maximum position i over 51 positions, obtained by rounding an equally spaced grid from 1 to n. At each position, we draw 1000 independent signal and random-threshold pairs. We compare the deterministic signal-stopping rule from Theorem 6 with the randomized policy from Theorem 4. The deterministic policy succeeds on $\mathcal{I}_i$ only when the signal reaches the maximum position, so its success probability depends strongly on i. The randomized

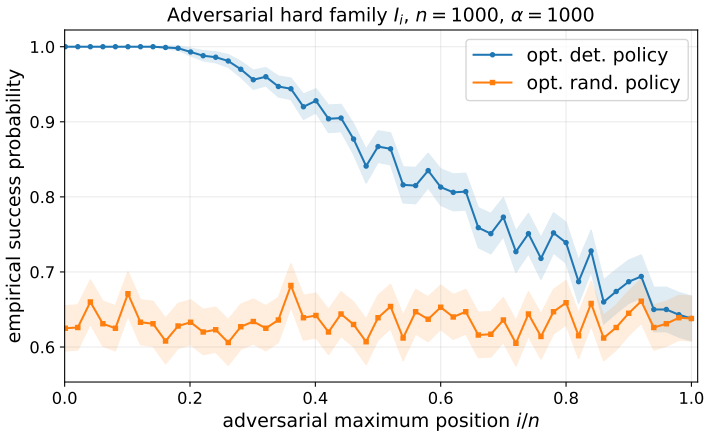

Figure 10: Adversarial position profile on the hard family $\mathcal{I}_i$. The deterministic signal-stopping rule is sensitive to the adversarial maximum position, while the randomized minimax policy from Theorem 4 exactly equalizes success in theory; the plotted Monte Carlo estimates fluctuate around this common value. Shaded bands are approximate pointwise 95% confidence intervals.

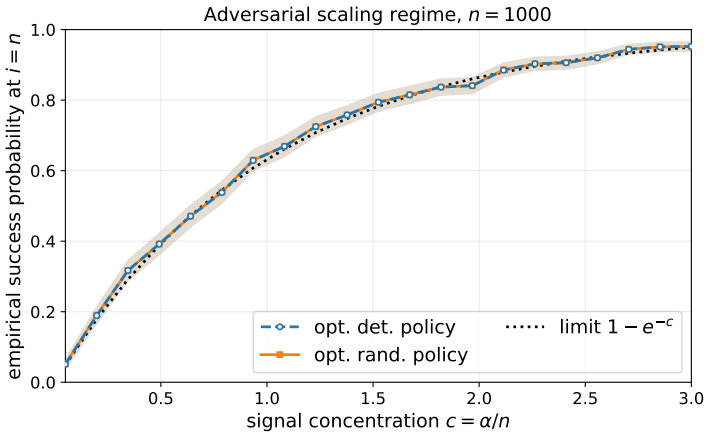

Figure 11: Adversarial scaling at $n = 1000$. Monte Carlo estimates on $\mathcal{I}_n$ are close to the limiting curve $1 - e^{-c}$. This instance is worst-case for the deterministic policy and representative of the exactly equalized randomized policy. The dashed deterministic and solid randomized curves nearly overlap because their exact finite-n gap is below 10^{-3} over the plotted range. Shaded bands are approximate pointwise 95% confidence intervals.

policy exactly equalizes success across the hard-family instances; the empirical estimates fluctuate around this common value.

Scaling regime $\alpha = cn$. The second adversarial experiment studies the regime $\alpha = cn$ at $n = 1000$, using 21 equally spaced values of c from 0.05 to 3. For each c, we estimate both policies on $\mathcal{I}_n$ using 1000 independent signal and random-threshold pairs. This is a worst-case instance for the deterministic policy, while the randomized policy exactly equalizes the hard-family instances. The exact deterministic and randomized values are given by Theorems 4 and 6, and Corollaries 5 and 7 shows that both approach $1 - e^{-c}$ as $n \to \infty$. Figure 11 compares the finite-n estimates with this limiting curve. The two policy curves nearly overlap at this scale, reflecting the vanishing benefit of randomization in the linear regime.

Full history versus the last signal for $m = 2$. The final adversarial experiment considers the multiple-signal model from Section 5. Revealing only the last signal $L = \max_j S_j$ is equivalent

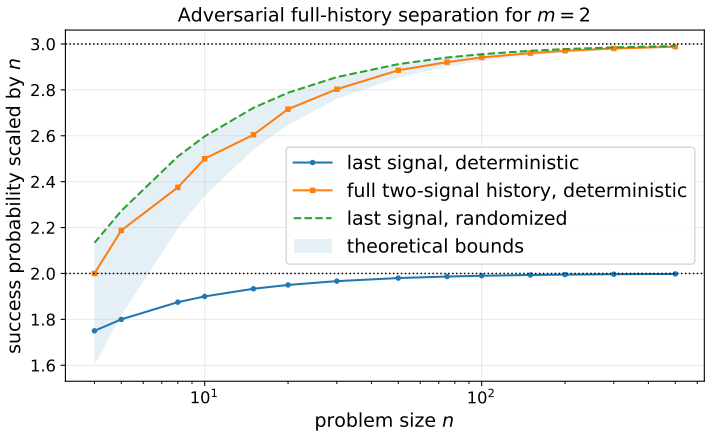

Figure 12: Full-history separation for $m = 2$ in the adversarial-order model. The full-history deterministic value from Theorem 18 lies between the bounds in Corollary 19 and is strictly larger than the deterministic last-signal value from Theorem 6.

to a single α-power signal with $\alpha = m$. Thus, for $m = 2$, the deterministic last-signal baseline is given by Theorem 6 with $\alpha = 2$, while the randomized last-signal value is given by Theorem 4; Theorem 8 shows that randomized algorithms gain no additional worst-case power from observing the full history. For deterministic algorithms, however, full history can help. We compute the full-history value for $m = 2$ numerically from the exact characterization in Theorem 18 and compare it with the last-signal baselines. The shaded region in Figure 12 shows the finite-n bounds from Corollary 19. The figure illustrates the separation: the deterministic full-history value scales like $\frac{3}{n}$, whereas the deterministic last-signal value scales like $\frac{2}{n}$.