Value of Learning in Online Decision Making: When does being Bayesian help?

Alireza AmaniHamedani
London Business School

Ali Aouad
Massachusetts Institute of Technology

Senem Işık
Stanford University

Amin Saberi
Stanford University

1. INTRODUCTION

Online decision-making under uncertainty raises a fundamental question about the relationship between *learning* and *decision-making*: when an algorithm begins interacting with an environment, should it rely only on its *prior knowledge* (e.g., training data), rely only on *online samples*, or combine the two in a *Bayesian* way? Although "be Bayesian" is often good advice, across online problems, the value of prior knowledge versus new data varies dramatically, and in some settings, one of these sources of information is redundant.

This tension is especially sharp in *optimal stopping* (and related posted-price settings), where classical frameworks can obscure the value of learning. When the distribution is known (prophet setting), there is nothing to learn. For an unknown i.i.d. distribution, [6] showed that, in the worst case, learning from early samples still cannot improve on the benchmark of $1/e$ for when samples are exchangeable but completely adversarial otherwise (secretary setting).

These observations motivate studying online decision problems in distributional models in which learning can genuinely improve decisions. We therefore introduce the *Mixture of Worlds* model, which captures a common form of structured uncertainty. In many stopping applications—pricing, timing a sale, recruiting, and ad auctions—one may know (from domain knowledge or past experiments) a small collection of plausible generative behaviors, but not which one of these behaviors is present in the current deployment. Early samples are then useful because they help identify the underlying distribution, while the stopping limits how long one can learn before making a decision.

Using this perspective, we distinguish three regimes across canonical online problems: ski rental, paging, and optimal stopping. Each problem reveals a distinct characterization of the value of different types of information, quantified by the marginal improvement in competitive ratio achievable by the best polynomial-time algorithm when given access to only prior knowledge, only online data, or both.

2. OUR MODEL: MIXTURE OF WORLDS

We first instantiate the model in *optimal stopping*: A decision-maker observes n values sequentially and must irrevocably stop at one of them, receiving the selected value. The benchmark is the offline optimum, which sees the entire sequence in advance and selects its maximum.

Learning-Augmented Algorithms: Theory and Applications (LATA) 2026
Ann Arbor, MI, USA

Under our model, there are k known candidate distributions to the decision-maker

$$\mathcal{P}_1, \ldots, \mathcal{P}_k, \qquad \text{and a known, correct prior } \pi \in \Delta_k,$$

over these candidates. Nature draws a *hidden* distribution $I \sim \pi$, and conditioned on I, the decison-makers observes

$$X_t \overset{\text{i.i.d.}}{\sim} \mathcal{P}_I, \qquad t = 1, \ldots, n.$$

Our quantitative bound depends on the support size k.

Related Work: Classical models of online algorithms sit at two opposite extremes: Adversarial analysis provides robust guarantees, but can be overly pessimistic; known-distribution or i.i.d. models yield sharper guarantees, but can be overly optimistic. Beyond-worst-case analysis seeks meaningful middle grounds between these extremes, where algorithms have access to imperfect but useful information, such as predictions [1, 13, 14, 17], priors [1, 3, 12, 9], samples [5, 11, 16], or prediction portfolios [2, 8, 10]. Our model introduces a Bayesian perspective on this landscape and interpolates smoothly between the classical known-and-correct distributional model and settings with no or many competing distributional assumptions, where predictions become increasingly noisy or inaccurate.

2.1 Ski Rental: Value of the Prior

Consider a person who decides to go skiing for an unknown number of days $T \in [n]$. Each day she chooses to *rent* for cost r or *buy* once for cost B (after which no further cost is incurred). Here, we prove that the prior on the distribution of T fully determines the optimal buying policy, making no additional online learning necessary.

PROPOSITION 1. *Let $T \in [n]$ be the random horizon with an arbitrary prior law—a strictly more general assumption than our Mixture of Worlds model—and let $\mathcal{H}_t$ denote the information available at time t. Then, for every set $A \subseteq [n]$,*

$$\Pr(T \in A \mid \mathcal{H}_t, T \geq t) = \Pr(T \in A \mid T \geq t).$$

In particular, the Bayes-optimal ski-rental policy is uniquely determined by the marginal distribution of T, and hence depends only on the prior.

2.2 Paging: Value of Online Samples

Take a *cache* C of capacity B and a universe U of $N \gg B$ pages. A request sequence of pages $X_i \in U$ is revealed online. At time t, serving X_t costs 0 if it is already cached, i.e. $X_t \in C$; otherwise a page fault of cost 1 occurs, the algorithm may load X_t into C, and (if the cache is full)

evict a page. The goal is to find an eviction algorithm that minimises the total cost.

Here, when X_i's are sampled i.i.d., we prove the opposite holds: online samples alone can drive asymptotically optimal performance, making prior information on request distribution asymptotically dispensable. In particular, let $S_i^\star \in \arg\max_{S\subseteq U:\,|S|=B} \mathcal{P}_i(S)$ denote the best static cache of size B under instance i. When $\sum_{i=1}^k \pi_i\big(1-\mathcal{P}_i(S_i^\star)\big) = \omega(n^{-1/4})$,[1] a simple algorithm that behaves like the well-known **LFU** (Least Frequently Used) policy based only on online samples achieves a competitive ratio $1+o(1)$.[2]

PROPOSITION 2. *Define the online algorithm* **ALG** *as follows. During the first $m_n := \lceil n^{1/2}\rceil$ requests,* **ALG** *serves requests using an arbitrary eviction rule. Let $\hat{\mathcal{P}}_n$ be the empirical distribution of these m_n requests, and let $\hat{S}_n \in \arg\max_{S\subseteq U:\,|S|=B} \hat{\mathcal{P}}_n(S)$. For every later fault,* **ALG** *evicts a page outside $\hat{S}_n$ whenever one is present. Then, $\mathbb{E}[\mathbf{ALG}] \le \mathbb{E}[\mathbf{OPT}^{\mathcal{P}}] + O(n^{3/4})$, where $\mathbf{OPT}^{\mathcal{P}}$ is the optimum policy that knows underlying distribution.*[3]

3. OPTIMAL STOPPING: VALUE OF BEING BAYESIAN

For *optimal stopping* with i.i.d. samples, neither extreme seems satisfactory. On the one hand, online learning alone cannot, in general, overcome the worst-case $1/e$ barrier [6, Theorem 2]. On the other hand, we show that a broad class of natural algorithms that rely only on the prior alone can perform arbitrarily poorly, including static-threshold policies with random tie-breaking, which gives $\big(1-\frac{1}{e}\big)$ approximation to optimum offline in the prophet setting:

PROPOSITION 3. *For any $\varepsilon > 0$, threshold τ, and $p \in [0,1]$, there exists a prior distribution such that the following policy has competitive ratio at most $O(\varepsilon)$ against the offline optimum: always accept if $X_t > \tau$, accept X_t with probability p if $X_t = \tau$, and reject otherwise.*

This shows that our setting is strictly harder. It also raises a computational question: must one compute the Bayes-optimal policy—often intractable[4] (indeed, PSPACE-hard in general [15, Theorem 6])—or can *approximately Bayesian* algorithms achieve near-optimal performance efficiently? Toward this goal, we give a Polynomial-Time Approximation Scheme (PTAS) that approximates the Bayes-optimal online policy, that is, the Bayesian dynamic program, to arbitrary precision.

THEOREM 1. *Assuming $\{X_i\}_{i=1}^n \subseteq [u,U]$,[5] for any target accuracy $\varepsilon \in (0,1)$, there is an online policy computable by a dynamic program with $O\left(\left(\frac{Cn^2}{\varepsilon}\log\left(\frac{nUk}{\varepsilon u}\right)\right)^k\right)$ states (and hence the same time complexity), for some absolute constant C, whose value-to-go at every time t is a $(1\pm\varepsilon)$-multiplicative approximation of the value-to-go of the Bayes-optimal stopping policy.*

In addition, for every fixed support size k, we provide a simple polynomial time algorithm—an explicit modification of the classical secretary algorithm—that beats the $1/e$ barrier by an additive term that decays exponentially in k. Thus, we answer in the affirmative, and in a strong quantitative sense, an open question of Correa et al. [7]: better-than-$1/e$ guarantee is indeed possible when the unknown i.i.d. distribution is drawn from a known prior supported on finitely many candidate distributions. Moreover, our result shows that the $1/e$ benchmark of [7, Theorem 5.1] is asymptotically correct only in the limit of *infinite* prior support.

THEOREM 2. *There exist absolute constants $c > 0$ and $C > 2$ such that for every fixed $k \ge 1$ and every instance $(\pi, \{\mathcal{P}_h\}_{h=1}^k)$, there exists a polynomial time algorithm* **ALG** *scuh that its competitive ratio satisfies*

$$\frac{\mathbb{E}[\mathbf{ALG}]}{\mathbb{E}[\max_i X_i]} \;\ge\; \frac{1}{e} + \frac{c}{C^k},$$

for all $n \ge n_0$ for some $n_0 \in \mathbb{N}$.

The key ingredient is a doubling-type argument that identifies a dominant distribution in the support of the prior whose contribution to the offline benchmark is large relative to the rest of the support. The policy is designed to combine two effects: it obtains a strict gain on the dominant distribution, while remaining close to the classical $1/e$ benchmark on the other relevant distributions. Finally, we show that our lower bound is asymptotically tight.

THEOREM 3. *For every $k \ge 1$ and every $\delta \in (0,1)$, there exists an instance $(\pi, \{\mathcal{P}_h\}_{h=1}^k)$ and $N_0(\delta) \in \mathbb{N}$ such that for all $n \ge N_0(\delta)$, the Bayes-optimal stopping time $\sigma^\star$ satisfies*

$$\frac{\mathbb{E}[X_{\sigma^\star}]}{\mathbb{E}[\max_i X_i]} \le \frac{1}{e} + \delta + O\left(\frac{n\log(1/\delta) + n\log n}{k}\right).$$

Besides making the dependence on n and k explicit, this yields the first *constructive* proof of the existential result of Correa et al. showing that one cannot beat $1/e$ in the Bayesian setting in the worst case [7, Theorem 5.1]. Our construction relies on three ideas. First, we stay close to the hard instance of [6, Theorem 2], so that the Bayes-optimal policy reduces to accepting only values that are running maxima. Second, following [1, Theorem 3.1], we impose a nested prior structure. This greatly simplifies the Bayesian state space. Finally, we show that the optimal dynamic program exhibits an asymptotic secretary-type threshold behavior as $k \to \infty$: it rejects almost all running maxima before about time n/e, and accepts almost all running maxima thereafter.

3.1 Future Work

We conjecture that the competitive ratio of the Bayes-optimal algorithm for optimal stopping is $\frac{1}{e} + \Theta\big(\frac{1}{k}\big)$ for any instance $(\pi, \{\mathcal{P}_h\}_{h=1}^k)$. Note that the algorithm in Theorem 2 does not use online samples beyond the calculation of the running maxima. Using the posterior more directly is an interesting direction for future work. One promising idea is to study *myopic* algorithms: policies that use the posterior at time t but ignore the fact that continuing at time t changes future beliefs endogenously.

[1] This assumption is mild and holds trivially for when the users request a page outside of the best cache with constant probability.
[2] The algorithm stated below keeps exact empirical frequencies during the learning phase. More memory-efficient implementations can replace this with streaming heavy-hitter methods such as [4], without changing the qualitative conclusion.
[3] This is a strictly harder benchmark than bayes-optimal policy.
[4] The main hardness is that there are uncountably many ways the posterior can evolve in the future, and each such trajectory affects the value-to-go and, hence, the current decision.
[5] This assumption is common in the literature (see [9]).

4. REFERENCES

[1] T. Bai, Z. Huang, C. S. Lee, and D. Li. Optimal stopping with a predicted prior, 2025.

[2] M.-F. Balcan, T. Sandholm, and E. Vitercik. Generalization in portfolio-based algorithm selection. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, pages 12225–12232. AAAI Press, 2021. Virtual Event, February 2–9, 2021.

[3] K. Banihashem, X. Chen, M. Hajiaghayi, S. Kim, K. Mahadik, R. Rossi, and T. Yu. Pandora with inaccurate priors, 2025.

[4] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. page 693–703, 2002.

[5] J. Correa, A. Cristi, B. Epstein, and J. A. Soto. Sample-driven optimal stopping: From the secretary problem to the i.i.d. prophet inequality. *Mathematics of Operations Research*, 49(1):441–475, Feb. 2024.

[6] J. Correa, P. Dütting, F. Fischer, and K. Schewior. Prophet inequalities for i.i.d. random variables from an unknown distribution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, page 3–17, New York, NY, USA, 2019. Association for Computing Machinery.

[7] J. Correa, P. Dütting, F. A. Fischer, K. Schewior, and B. Ziliotto. Unknown I.I.D. prophets: Better bounds, streaming algorithms, and a new impossibility (extended abstract). In J. R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, Virtual Conference, January 6-8, 2021*, LIPIcs, pages 86:1–86:1. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[8] M. Dinitz, S. Im, T. Lavastida, B. Moseley, and S. Vassilvitskii. Algorithms with prediction portfolios. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[9] P. Dütting and T. Kesselheim. Posted pricing and prophet inequalities with inaccurate priors. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, page 111–129, New York, NY, USA, 2019. Association for Computing Machinery.

[10] M. Eliáš, H. Kaplan, Y. Mansour, and S. Moran. Learning-augmented algorithms with explicit predictors. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc.

[11] H. Kaplan, D. Naori, and D. Raz. Competitive analysis with a sample and the secretary problem, 2019.

[12] G. Kehne and T. Kesselheim. Prophet and secretary at the same time, 2025.

[13] T. Lykouris and S. Vassilvitskii. Competitive caching with machine learned advice. *Journal of the ACM*, 68(4), July 2021.

[14] M. Mahdian, H. Nazerzadeh, and A. Saberi. Allocating online advertisement space with unreliable estimates. In *Proceedings of the 8th ACM Conference on Electronic Commerce*, Proceedings of the 2007 ACM Conference on Economics and Computation, page 288–294, New York, NY, USA, 2007. Association for Computing Machinery.

[15] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of markov decision processes. *Mathematics of Operations Research*, 12(3):441–450, Aug. 1987.

[16] A. Rubinstein, J. Z. Wang, and S. M. Weinberg. Optimal single-choice prophet inequalities from samples. 2025.

[17] J. H. Shen, E. Vitercik, and A. Wikum. Algorithms with calibrated machine learning predictions. In *Forty-second International Conference on Machine Learning*, 2025.