Prophets, Secretaries, and Maximizing the Probability of Choosing the Best

Hossein Esfandiari Google Research
MohammadTaghi HajiAghayi University of Maryland
Brendan Lucier Microsoft Research
Michael Mitzenmacher Harvard University

Abstract

Suppose a customer is faced with a sequence of fluctuating prices, such as for airfare or a product sold by a large online retailer. Given distributional information about what price they might face each day, how should they choose when to purchase in order to maximize the likelihood of getting the best price in retrospect? This is related to the classical secretary problem, but with values drawn from known distributions. In their pioneering work, Gilbert and Mosteller [*J. Amer. Statist. Assoc. 1966*] showed that when the values are drawn i.i.d., there is a thresholding algorithm that selects the best value with probability approximately 0.5801. However, the more general problem with non-identical distributions has remained unsolved.

In this paper we provide an algorithm for the case of non-identical distributions that selects the maximum element with probability $1/e$, and we show that this is tight. We further show that if the observations arrive in a random order, this barrier of $1/e$ can be broken using a static threshold algorithm, and we show that our success probability is the best possible for any single-threshold algorithm under random observation order. Moreover, we prove that one can achieve a strictly better success probability using more general multi-threshold algorithms, unlike the non-random-order case. Along the way, we show that the best achievable success probability for the random-order case matches that of the i.i.d. case, which is approximately 0.5801, under a "no-superstars" condition that no single distribution is very likely ex ante to generate the maximum value. We also extend our results to the problem of selecting one of the k best values.

1 Introduction

Suppose we are given a sequence of real numbers one by one, drawn from independent but not necessarily identical distributions known in advance. We can keep a single number from the sequence, but this choice must be made online. At each observation, we can either select the current number or push our luck and continue to the next observation. Our goal is to maximize the probability of selecting the maximum (or equivalently minimum) number from the sequence.

As a toy application, consider an airfare platform that provides a service of suggesting when a buyer should purchase their ticket for the lowest fare. Such a platform has distributional information about how expensive the fare will be each day before the flight. Users hope to avoid the regret of purchasing at a suboptimal price, and this incentivizes the platform to maximize the likelihood of suggesting the best price in hindsight. Given a model that maps time-before-flight and other fixed information (such as location and airline) to a distribution over prices, how should the platform make its online recommendations, and how likely is it to achieve the best price?

This question is related to the classical *secretary problem*. In the secretary problem, we receive a sequence of randomly permuted numbers 1 to n in an online fashion. We are given the numbers one by one, but each time we observe a number, we see only its relative rank compared to the previously observed numbers. At each observation, we have the option to stop the process and select the most recent number. The goal is to maximize the probability of selecting the maximum (or equivalently minimum) number. For this problem, Dynkin 1963 presents a simple but elegant algorithm that succeeds with probability at least $1/e$; indeed, the success probability converges from above

Proceedings of the 23[rd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

to $1/e$ as n grows large, and $1/e$ is the best possible bound (up to lower order terms) we can achieve for this problem.

A natural variation of the problem assumes that the numbers are drawn from the same known distribution, and the numbers themselves are revealed one by one. In their classic work, Gilbert and Mosteller 1966 consider this so-called "full information" case.[1] As a starting point, they show that one can pick a single threshold τ such that stopping at the first value larger than τ will select the maximum value with probability approximately 0.517 (asymptotically as n grows large). For the general case where one can use a distinct threshold at each step, they show that with the appropriate choice of thresholds one succeeds at stopping at the maximum value with probability approximately 0.5801 (again, asymptotically as n grows large; both bounds are tight). These bounds significantly improve upon the $1/e \approx 0.37$ result for the secretary problem, which corresponds to the setting where the underlying distribution is not known and only the relative ranks are obtained.[2]

Since the work of Gilbert and Mosteller, there has been a vast literature on secretary problems going well beyond the scope of this paper. We refer the interested reader to a survey by Freeman for an overview of this activity from the perspective of stopping theory (Freeman, 1983). The full information case has received less attention, but there has been a notable line of work considering variations such as n being randomized (Porosiński, 1987) and/or it being possible to revisit previously-observed values with a probability of failure (Petruccelli, 1981). To our knowledge, this literature on the full information case has focused exclusively on the case of i.i.d. values.

Our Contributions We consider the more general problem of selecting the maximum (or minimum) value when the numbers are drawn from distributions that are independent but *not necessarily identical*. Our first result is that there is an algorithm that achieves a success probability of $1/e$ in this non-i.i.d. setting, matching the original secretary problem. This is tight up to lower-order terms. Our algorithm uses a single fixed threshold rule, and thus applies even if the values are revealed in an adaptively adversarial order. Our lower bound holds even if the order is known in advance and applies to arbitrary algorithms, showing that a simple fixed-threshold rule is asymptotically optimal. This expands the long-standing result for the i.i.d. case due to Gilbert and Mosteller 1966 to the setting with different distributions.

We next consider a random-order model, where the values are drawn from arbitrary independent distributions but are presented in a uniformly random order. The i.i.d. setting of Gilbert and Mosteller 1966 is a special case of this random order setting, where all observed values are chosen from the same distribution. Our second result generalizes the result of Gilbert and Mosteller to show that in the random-order setting, it is possible to select the maximum value with probability at least 0.517, using a single-threshold algorithm. This improves on the adversarial-order setting, and matches the tight bound for single-threshold algorithms for the i.i.d. case (Gilbert and Mosteller, 1966).

Still in the random-order model, we next present an algorithm that breaks this barrier of 0.517 using multiple thresholds. As a corollary, algorithms that use a single threshold are not optimal in the random-order model. Our approach is to consider a natural "no-superstars" condition, which is that no single distribution has more than a certain constant probability (ex ante) of generating the maximum value. This captures scenarios where no single entry has a non-vanishing impact on the problem's solution in the limit as the problem size n grows large.[3] We show that under such an assumption, there is an algorithm that succeeds with probability arbitrarily close to 0.5801, the tight success probability obtainable in the i.i.d. setting with multiple thresholds, in the limit as n grows large. If the no-superstars assumption is violated, the presence of a highly dominant distribution again makes it possible to improve over the single-threshold bound of 0.517.

It is natural to compare these results with the *prophet inequality*, introduced by Krengel et al. 1978; 1977. In the prophet inequality problem, the goal is to maximize the expected value of the number selected rather than the probability of selecting the maximum. The classic prophet inequality is that one can achieve half of the expected maximum value using a single threshold algorithm, and this is tight. The *prophet secretary* model (Esfandiari et al., 2017) considers this goal of maximizing expected value in the random-order model, which admits improved results. One can view our results as extending classic "secretary-style" results for best-choice problems to settings typical of prophet inequalities, with independent but non-

[1]Despite the name, this is similar to the "incomplete information" setting in mechanism design. We will use the term "full information" in the sense of secretary problems, meaning that the distributions are known in advance.

[2]Interestingly, one of the original motivations Gilbert and Mosteller 1966 provide is a simplified model for an atomic bomb inspection program, which may have arisen from Mosteller's work in Samuel Wilks's Statistical Research Group in New York city during World War II on statistical questions about airborne bombing.

[3]We can think of this as a *large market* condition.

identical distributions.

One distinction between the best-choice problem and a prophet inequality is that, in the best-choice problem, it is typically better to avoid a "non-robust" solution that achieves high expected value by accepting a very large number with very small probability, but otherwise does not obtain much value. Motivated by this connection to robustness, one might relax the desideratum of picking only the highest number, and aim instead to obtain one of the top few values with high probability. To this end, we consider a variant of our problem where the goal is to maximize the probability of selecting any of the top k values. A similar variant has been studied for the secretary problem by Gusein-Zade 1966, who shows that there is an algorithm whose failure probability is at most $O(\frac{\log k}{k})$. When values are drawn i.i.d. from a known distribution, Gilbert and Mosteller 1966 study the case $k = 2$ and solve for the limiting probability of success. We extend this to arbitrary k and arbitrary distributions presented in an adversarial order, and show that there is an algorithm with failure probability exponentially small in k. Moreover, this is the best possible bound, up to coefficients in the exponent, even in the i.i.d. setting.

As one of our main tools in our analysis, we use Le Cam's theorem (Le Cam, 1960), which (as we describe below) connects sums of Bernoulli random variables and discrete Poisson distributions. This result, along with coupling techniques and other additions for our setting, allow us to represent the probability distribution for the maximum (over several different distributions) by discrete Poisson distributions. This variation of a "Poissonization" argument for these settings appears novel, and may be of its own interest.

1.1 Results and Techniques

In what follows, we refer to the best-choice prophet inequality problem and best-choice prophet secretary problem for the variations we consider, where the goal is to maximize the probability of choosing the highest observed value given distributions presented in adversarial and random order, respectively. We start by obtaining a tight bound for the best-choice prophet inequalities problem: we provide an algorithm that selects the maximum with probability at least $\frac{1}{e}$ and show that there is no algorithm that selects the maximum with probability at least $\frac{1}{e} + \epsilon$ for any constant $\epsilon > 0$. Although the probability of success here is the same as for the classical secretary problem, the proof and corresponding algorithm are not the same. Our algorithm is based on choosing a suitable threshold and accepting any observation above that threshold. We choose the threshold to optimize the probability that exactly one element lies above it, since we are guaranteed to accept the largest value in this case. Perhaps surprisingly, our lower bound shows that this analysis is tight, even for an arbitrary selection rule with advance knowledge of the arrival order.

We next provide a single threshold algorithm for the best-choice prophet secretary problem that selects the maximum with probability at least 0.517. This result utilizes some of the technology used for the best-choice prophet inequality result. We also extend our analysis to the top-k-choice prophet inequality problem, and provide a single threshold algorithm that selects one of the top k values with probability at least $1 - e^{-c_1 k}$, where $c_1 > 0$ is a fixed constant. We also show that this exponential dependence on k is tight even in the i.i.d. setting: there is a constant $c_2 > 0$ such that no algorithm can select one of the top k values with probability greater than $1 - e^{-c_2 k}$. This tightness result involves arguing that an arbitrary algorithm must become "trapped" at some point in the observation sequence, with at least an exponential probability; conditional on what it has seen, there is a non-negligible chance that all of the top k values have already been observed, but also a non-negligible chance that all of them are yet to come.

All of the algorithms above use a single fixed threshold. For the best-choice prophet inequality problem our lower bound shows that single-threshold algorithms achieve tight results, but for the best-choice prophet secretary problem we show that this is not the case. Designing and analyzing multiple-threshold algorithms is significantly more challenging, as dependencies and correlations naturally arise. To overcome this, we develop an alternative approach for analyzing the setting of multiple distributions in a random order. The intuition is that for a large number of observations n, we can split the observations into consecutive groups of size n/T for a suitable constant T, such that we can think of the maximum of each group as being approximately from an i.i.d. distribution corresponding to a sample of n/T distributions from the n overall distributions. That is, each group of n/T distributions is sufficiently similar that we can view the problem as very similar to the best-choice problem for T i.i.d. observations. Formalizing this closeness allows us to nearly achieve the same worst-case performance of a T-threshold scheme in the i.i.d. setting. This result requires a technical "no-superstars" condition, which is that the *a priori* probability of any specific distribution being the maximum is $o(1)$. Using this technique, and under no-superstars assumption, we design a threshold-based algorithm whose success probability converges to 0.5801 as n grows large, which is tight even for the i.i.d. setting. On the other hand, we show that if the no-superstars assumption is violated and there exists a distribution that has more than a certain constant probability of generating the maximum value as n grows large, one can improve the single-threshold analysis. Combining these methods leads to an unconditional improvement over the optimal worst-case bound for single-threshold algorithms.

We briefly note that, unlike the expectation version of prophet inequalities (Esfandiari et al., 2017), in this setting of best-choice prophet inequalities, all our results trivially extend to the setting where we want to maximize the probability of finding the minimum element as well.

1.2 Poissonization Technique

One approach used in (Gilbert and Mosteller, 1966) involves setting a threshold and considering the number of observations above that threshold. In the case of i.i.d. distributions for the observations, this number is the sum of i.i.d. Bernoulli random variables, which is known to converge to a Poisson distribution in the setting we consider (where the expected number of positive observations is constant as the number of observations grows large).

A helpful tool in extending such results to the setting where distributions may differ for observations is Le Cam's theorem (Le Cam, 1960). The basic statement of Le Cam's Theorem is the following: let $X_1, \cdots, X_n$ be a sequence of Bernoulli random variables where $\Pr[X_i = 1] = p_i$ and $\lambda = \sum_{i=1}^{N} p_i$. We have

$$\sum_{k=0}^{\infty} \left| \Pr\left[\sum_{i=1}^{n} \mathcal{X}_i = k\right] - \frac{\lambda^k e^{-\lambda}}{k!} \right| < 2 \sum_{i=1}^{N} {p_i}^2.$$

Intuitively, Le Cam's Theorem says that when the probability of each random variable being 1 in a sequence of Bernoulli random variables is sufficiently small (e.g. $O(\frac{1}{N})$), the sum is well approximated by a Poisson distribution. There are a number of interesting proofs of Le Cam's Theorem (see the survey (Steele, 1994)), including proofs that slightly improve the constant on the right hand side, but this general bound suffices for our purposes.

2 Further Related Work

Starting with the work of Dynkin 1963, there has been a long line of research on variants of the secretary problem. See the survey by Ferguson 1989 for a lighthearted but thorough historical treatment, and the review paper by Freeman 1983 for many generalizations.

There have likewise been many generalizations of the prophet inequality, since the initial work of Garling, Krengel, and Sucheston (1978; 1977). One of the first generalizations was the *multiple-choice prophet inequality* (Kennedy, 1987; Kennedy et al., 1985; Kertz, 1986) in which we are allowed to pick k items and the goal is to maximize their sum. Alaei (2014) gives an almost tight $(1 - 1/\sqrt{k+3})$-approximation algorithm for this problem (the lower bound is due to Hajiaghayi et al. (2007)), where the *approximation factor* is the ratio of the expectation of the algorithm to the expectation of the optimum. Similarly, the *multiple-choice secretary* problem was first studied by Hajiaghayi et al. (2004), and Kleinberg (2005) gives a $(1 - O(\sqrt{1/k}))$-approximation algorithm.

Other than Dynkin (1963), generally follow-up work considers approximation factors instead of maximizing the probability of obtaining the best. An interesting exception is ?, who provides a general approach for determining the optimal stopping time for choosing the maximum of a sequence of i.i.d. random varaibles (along with approaches for finding the optimal stopping time for some related problems). This work does not determine bounds on the probability of choosing the maximum, as we do here for the problems we consider.

The research investigating the relation between prophet inequalities and online auctions is initiated in Hajiaghayi et al. (2007); Chawla et al. (2010). This lead to several interesting follow up works for matroids Yan (2011) and matchings Alaei et al. (2012). Meanwhile, the connection between secretary problems and online auctions is first explored in Hajiaghayi et al. Hajiaghayi et al. (2004). Its generalization to matroids is considered in Babaioff et al. (2007); Lachish (2014); Feldman et al. (2015) and to matchings in Goel and Mehta (2008); Korula and Pál (2009); Mahdian and Yan (2011); Karande et al. (2011); Kesselheim et al. (2013); Guruganesh and Singla (2017).

In the prophet secretary model, Esfandiari et al. Esfandiari et al. (2017) give a $(1-1/e)$-approximation in the special case of a single item. Going beyond $1-1/e$ has been challenging. Only recently, Abolhasani et al. Abolhassani et al. (2017) and Correa et al. Correa et al. (2017) improve this factor for the single item i.i.d. setting. Very recently, Ehsani et al.Ehsani et al. (2018) extend prophet secretary for combinatorial auctions and matroids as well.

3 Notation

In the *best-choice prophet inequality problem*, we are given a set of distributions $\{D_1, \ldots, D_n\}$. We then observe an online sequence of values $x_1, \cdots, x_n$, where each x_i is drawn independently from D_i, presented in an arbitrary order. When value x_i is observed, we must irrevocably decide whether or not to choose that value. Once we choose a value, the process stops. A value that has been observed but not chosen cannot be chosen later. The goal is to maximize the probability that the value chosen is equal to $\max_i\{x_i\}$. We emphasize that the order in which the values are presented is arbitrary and not known in advance. We refer to the case with identical distributions as the *i.i.d. setting.*

The *best-choice prophet secretary problem* is identical, except that the values are presented in a uniformly random order. That is, after applying a random permutation $\Pi = \langle \pi_1, \ldots, \pi_n \rangle$ on the sequence of x_i values, they are presented in that order, so that at step k, π_k and x_{π_k} are revealed. Again, the goal is to maximize the probability of choosing a maximum value.

Our algorithms are threshold-based, where we choose a value if and only if it lies above a suitable threshold. We use $\mathcal{T} = \langle \tau_1, \ldots, \tau_n \rangle$ to refer to a sequence of thresholds; thus, we check for example whether $x_{\pi_k} \geq \tau_k$. In the case that $\tau_1 = \tau_2 = \ldots = \tau_n = \tau$, we say that the algorithm is a *single-threshold algorithm.*

In our proofs, we will assume for notational convenience that the distributions are *atomless:* the probability distributions are continuous, so that no single value takes on a non-zero probability. We use this assumption only to define the inverse of a given cumulative distribution; i.e., to find a value τ such that $\Pr_{x \sim D}[x \geq \tau] = p$ for some fixed $p \in [0, 1]$. This is only for convenience, and our results actually apply to the general case with atoms, using the following reduction based on using an auxiliary random number to break ties (which we believe is folklore). If there exists a value τ such that $\Pr_{x \sim D}[x \geq \tau] > p$ but also $\Pr_{x \sim D}[x \leq \tau] \geq p$ (i.e., there is an atom that prevents the desired inversion), then we can modify our random process to include a random variable y drawn from the uniform distribution on $[0, 1]$, and augment threshold τ with a secondary threshold $\bar{y}$. We will then interpret the event $[x \geq \tau]$ to mean $[(x > \tau) \vee ((x = \tau) \wedge (y \geq \bar{y}))]$, and set y so that, under this definition, $\Pr_{x \sim D}[x \leq \tau] = p$. With this reduction in mind, we will assume throughout that distributions are atomless without further comment.

4 Best-Choice Algorithms with a Single threshold

In this section, we describe algorithms and lower bounds for the best-choice prophet inequality problem (in Section 4.1) and the best-choice prophet secretary problem (in Section 4.2). All of the algorithms in this section will be single-threshold algorithms.

4.1 Best-Choice Prophet Inequalities

First we show that it is possible to choose the maximum value with probability at least $\frac{1}{e}$, using a single threshold, for the best-choice prophet inequality problem.

Theorem 1 *For the best-choice prophet inequality problem, there is an algorithm that succeeds with probability at least* $\frac{1}{e}$.

Proof : We will warm up by proving an easier result: a simple single-threshold algorithm that succeeds with probability $1/4$. We'll then show how to improve this to $1/e$. Our algorithm will select threshold τ such that $\Pr[\max_{i=1}^n(x_i) \geq \tau] = 1/2$, and choose the first value that is at least τ. From the definition of τ, the algorithm chooses a value with probability $1/2$, otherwise it chooses nothing. Conditional on having chosen a value, the algorithm will certainly succeed if no subsequent value is strictly greater than τ. But the probability of a subsequent value lying above τ is at most $1/2$, the probability that *any* of the n observations is greater than τ. So the probability of success, conditional on having selected an item, is at least $1/2$, leading to a total success probability of at least $1/4$.[4]

We can modify the algorithm above to improve the success probability to $1/e$. Namely, the algorithm will set threshold τ so that $\Pr[\max_{i=1}^n(x_i) \leq \tau] = 1/e$ and pick the first number that is larger than τ. We show that with probability at least $1/e$ there is exactly one number which is larger than τ, which implies the desired result. Let $p_i = \Pr[x_i > \tau]$. By the way we choose τ, we have $\prod_{i=1}^n(1 - p_i) = 1/e$.

We now consider the probability that exactly one number is larger than τ, and show that it is at least $1/e$; this completes the proof.[5] The probability that the jth observed value is larger than τ but all others are not is $\frac{p_j}{1-p_j}\prod_{i=1}^n(1 - p_i) = \frac{1}{e}\frac{p_j}{1-p_j}$. We briefly note the fact that $e^x \geq 1 + x$ implies (using $x = p_j/(1 - p_j)$)

$$\frac{p_j}{1 - p_j} \geq \ln\left(\frac{1}{1 - p_j}\right). \tag{1}$$

Now the probability that exactly one number is larger than τ is

$$\frac{1}{e}\sum_{j=1}^n \frac{p_j}{1 - p_j} \geq \frac{1}{e}\sum_{j=1}^n \ln\left(\frac{1}{1 - p_j}\right) = \frac{1}{e}\ln\frac{1}{\prod_{j=1}^n(1 - p_j)} = \frac{1}{e}.$$

[4] This warm up is similar to ?.

[5] We thank an anonymous reviewer for providing us this simplification over our prior proof.

Here the first line follows from Inequality 1, and the last line from $\prod_{j=1}^{n}(1-p_j) = 1/e$. $\square$

Our algorithm uses only a single fixed threshold as its stopping rule. One might suspect that a more complicated algorithm, perhaps one that modifies its thresholds adaptively or employs randomization, would perform better. Our next result is that this is not the case: no online algorithm can guarantee a success probability strictly better than $\frac{1}{e}$. We provide the proof of this theorem in Appendix B

Theorem 2 *For any constant $\varepsilon > 0$, there is no algorithm that succeeds with probability $\frac{1}{e} + \varepsilon$ for the best-choice prophet inequality problem.*

4.2 Best-Choice Prophet Secretary

In this subsection we show a single threshold suffices to provide an algorithm that chooses the maximum value with probability 0.517 for best-choice prophet secretary. To begin, we provide a simple analysis that achieves this 0.517 probability for best-choice prophet inequalities with i.i.d. distributions. We note that this result was presented in Gilbert and Mosteller (1966), with the constant calculated numerically for large values of n. We essentially follow their argument, but provide a formal justification for their numerical results.

Theorem 3 *For sufficiently large n, there exists a single threshold algorithm that chooses the maximum value with probability arbitrarily close to $\max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right) \approx 0.5173$, for best-choice prophet inequalities with i.i.d. distributions, and this is tight for single-threshold algorithms.*

Proof : Let τ be given by $\Pr[\max_{i=1}^{n}(x_i) \le \tau] = \mathcal{P}$, and $p = \Pr[x_i \ge \tau] = 1 - \mathcal{P}^{1/n}$ for $\mathcal{P}$ to be given later. Let $\mathcal{K}$ be the random variable indicating the number of x_i that are greater than τ. When $\mathcal{K} \ge 1$, due to symmetry each of these $\mathcal{K}$ items is the maximum with probability $1/\mathcal{K}$, and since we pick the first item that is greater than τ, when $\mathcal{K} \ge 1$ the maximum is chosen with probability $1/\mathcal{K}$. So, we pick the maximum with probability at least

$$\sum_{k=1}^{n}\Big(\frac{1}{k}\Pr[\mathcal{K} = k]\Big).$$

Here $\mathcal{K}$ is sum of Bernoulli random variables, and so the probability we choose the maximum is simply

$$\sum_{k=1}^{n}\Big(\frac{1}{k}\binom{n}{k}p^k(1-p)^{n-k}\Big).$$

For large n we may use that the limit of the Bernoulli distribution becomes a Poisson distribution, and use numerical calculations and Le Cam's theorem to obtain the result. Specifically, take $\mathcal{P} = (1-1.501/n)^n \simeq e^{-1.501}$ and $p = \Pr[x_i \ge \tau] = 1.501/n$, where the 1.501 is determined numerically. By Le Cam's theorem $\sum_{k=0}^{\infty}\left|\Pr[\mathcal{K} = k] - \frac{\lambda^k e^{-\lambda}}{k!}\right| < 2np^2$, where $\lambda = np = 1.501$. This gives us $\sum_{k=1}^{\infty}\left|\frac{1}{k}\Pr[\mathcal{K} = k] - \frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right| < 2np^2 < \frac{6}{n}$. Therefore the probability that we pick the maximum is at least

$$\begin{aligned}\sum_{k=1}^{n}\Big(\frac{1}{k}\Pr[\mathcal{K} = k]\Big) &\ge \sum_{k=1}^{n}\Big(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\Big) - \frac{6}{n}\\ &\ge 0.5173 - \frac{6}{n} \ge 0.517,\end{aligned}$$

where the second inequality is calculated numerically for $\lambda = 1.501$ and the last inequality is by assuming $n \ge 20000$. We note that by taking n large enough, we can obtain a success probability arbitrarily close to the sum $\max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right)$ using the same argument. This is an asymptotic upper bound by a similar argument, so this success probability is tight. $\square$

We are now ready to extend Theorem 3 to the more general best-choice prophet secretary problem. Notice that the following theorem does not require n to be large, so even when applied to the special case of i.i.d. distributions it extends Theorem 3 to general n.

Theorem 4 *There exists a single threshold algorithm that chooses the maximum value with probability at least $\max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right) \approx 0.5173$, for the best-choice prophet secretary problem.*

Proof : As in Theorem 3, we set τ such that $\Pr[\max_{i=1}^{n}(x_i) \le \tau] \simeq e^{-1.501}$ and pick the first number which is at least τ. We clarify the exact value of τ later in the proof after we present the required notation. To analyze the algorithm, for some arbitrary small ε' we replace each distribution D_i with a bag of n^2/ε' identical and independent copies of a dummy distributions D'_i, where the distribution of the maximum of the n^2/ε' copies of D'_i is equivalent to D_i. We let x_i^j to be the realization of the j'th copy of D'_i, let $p_i^j = \Pr\left[x_i^j \ge \tau\right]$, and let $n' = n^3/\varepsilon'$ to be the total number of dummy distributions. By the way we have defined the dummy distributions, the distribution of the maximum of all dummy distributions is equivalent to the distribution of the maximum of the original problem.

The bags arrive in a random order and upon the arrival of each bag we observe the realization of the maximum number in the bag. The first time we face a bag with at least one number above the threshold, we stop and pick the maximum number in the bag. Again, the distribution of the value chosen in this framework is equivalent to that of our threshold algorithm on the actual distributions.

Let $\mathcal{K}$ be the random variable indicating the number of x_is that are greater than τ and let $\mathcal{K}'$ be the random variable indicating the number of x_i^js that are greater than τ. In fact, if for some i we have $x_i \geq \tau$, then for some j we have $x_i^j \geq \tau$. Hence we have $\mathcal{K}' \geq \mathcal{K}$. Notice that if $\mathcal{K}' \geq 1$ with probability $1/\mathcal{K}$ the bag that contains the maximum number arrives first and we select the maximum number; otherwise, we do not. Thus, we choose the maximum with probability

$$\sum_{k=1}^{n'} \Big(\Pr\left[\text{We choose the max} \middle| \mathcal{K}' = k\right] \Pr\left[\mathcal{K}' = k\right] \Big)$$

$$\geq \sum_{k=1}^{n'} \Big(\frac{1}{k} \Pr\left[\mathcal{K}' = k\right] \Big). \qquad \text{since } \mathcal{K}' \geq \mathcal{K} \qquad (2)$$

Now we are ready to set the value for τ given at the beginning of the proof; specifically, we set τ so that $\sum_{i=1}^{n} \sum_{j=1}^{n^2/\varepsilon'} p_i^j$ equals the value of λ that maximizes $\max_\lambda \sum_{k=1}^{\infty} \left(\frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!} \right)$, which is approximately 1.501. This corresponds to $\lambda = 1.501$ for Le Cam's Theorem. Also for any i and j we have $p_i^j \leq \varepsilon'/n^2$. Using Le Cam's Theorem we have

$$\sum_{k=1}^{\infty} \left| \frac{1}{k} \Pr\left[\mathcal{K}' = k\right] - \frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!} \right|$$

$$\leq \sum_{k=1}^{\infty} \left| \Pr\left[\mathcal{K}' = k\right] - \frac{\lambda^k e^{-\lambda}}{k!} \right|$$

$$< 2 \sum_{i=1}^{n} \sum_{j=1}^{n^2/\varepsilon'} {p_i^j}^2 \qquad \text{Le Cam's Theorem}$$

$$\leq 2 \frac{n^3}{\varepsilon'} \times \left(\frac{\varepsilon'}{n^2}\right)^2 \leq \varepsilon', \qquad \text{since } p_i^j \leq \frac{\varepsilon'}{n^2} \text{ and } n > 1.$$

This immediately gives us $\sum_{k=1}^{\infty} \left(\frac{1}{k} \Pr\left[\mathcal{K}' = k\right] \right) \geq \sum_{k=1}^{\infty} \left(\frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!} \right) - \varepsilon'$. Therefore, the probability that our algorithm picks the maximum is at least

$$\sum_{k=1}^{n'} \Big(\frac{1}{k} \Pr\left[\mathcal{K}' = k\right] \Big) = \sum_{k=1}^{\infty} \Big(\frac{1}{k} \Pr\left[\mathcal{K}' = k\right] \Big)$$

$$\geq \sum_{k=1}^{\infty} \Big(\frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!} \Big) - \varepsilon',$$

where the first equality holds since for $k > n'$, $\Pr\left[\mathcal{K}' = k\right] = 0$. Recall that ε' is an arbitrary small positive number and the algorithm does not depend on ε'. Hence, the probability that our algorithm picks the maximum is at least $\sum_{k=1}^{\infty} \left(\frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!} \right)$ as claimed. □

We note that since the lower bound in Theorem 4 matches the upper bound on the performance of any single-threshold algorithm from Theorem 3, we can conclude that the algorithm in Theorem 4 is best-possible among single-threshold algorithms for best-choice prophet secretary.

5 Top-k-Choice Algorithms

In this section we consider a variant of our best-choice problems, where the goal is relaxed to choosing one of the k largest values. Here $k > 1$ is fixed as n grows large. As before, we can make only a single choice; doing so stops the process, and that is the final selection. We first show that for the top-k-choice prophet inequality problem, where the distributions are presented in an arbitrary order, there is a single-threshold algorithm whose probability of failure is exponentially small in k.

Theorem 5 *For any $k \geq 1$, there exists an algorithm for the top-k-choice prophet inequality problem that succeeds with probability at least $1 - 2e^{-\gamma k}$, where $\gamma = (3 - \sqrt{5})/2$.*

The algorithm in Theorem 5 sets its threshold τ so that the expected number of values greater than τ is exactly γk. The result then follows by applying standard concentration bounds (Chernoff) to show that it is exponentially unlikely (in k) that no values are greater than τ, and also exponentially unlikely that strictly more than k values are greater than τ. The formal details are deferred to Appendix D.

One thing to note about the bound in Theorem 5 is that it is independent of n, which we can take to be very large relative to k. It's tempting to imagine that one could improve this error in special cases such as the i.i.d. setting. Our next result shows that this is not possible. One cannot do better than an exponentially decreasing error in k, even for the i.i.d. setting and hence also for the top-k-choice prophet secretary problem.

We note that for such a bound one cannot simply condition on observing a certain worst-case ordering over a collection of $\theta(k)$ distributions, as the probability of seeing any particular permutation of $\theta(k)$ elements is $e^{-\theta(k \log k)}$. The intuition of our proof is that, say halfway through the process, there is at least an exponentially small probability that the algorithm becomes "trapped:" given what it has seen, there is at least an exponentially small probability that all of the top k values were present in the first half, but also at least an exponentially small probability that all of the top k values appear in the second half. Thus, regardless of what the algorithm has done, an exponential error bound cannot be avoided. Formalizing this intuition takes some care.

Theorem 6 *There exists a constant c such that, for any fixed $k \geq 1$, no algorithm for the top-k-choice prophet inequality problem with identical distributions selects the maximum with probability more than $1 - e^{-c \cdot k}$.*

6 Improved Best-Choice Prophet Secretary with Multiple Thresholds

As we showed in Section 4.1, a single threshold algorithm achieves tight results for best-choice prophet inequalities. However, this does not seem to be true for best-choice prophet secretary. In this section, which captures our most technical result, we seek to go beyond the single threshold algorithms and design a more efficient algorithm for best-choice prophet secretary. Our algorithm will use multiple thresholds. First we provide an algorithm for inputs with an additional assumption that we call the *no-superstars assumption*, which is that no single observation has too large a probability, *a priori*, of being the largest value. Then we use this algorithm to provide an unconditional algorithm for best-choice prophet secretary that improves upon single threshold algorithms.

Definition 7 (No-Superstars Assumption.) *We say that a set of distributions $\{D_1, \ldots, D_n\}$ satisfies the* no ε-superstars assumption *if, for all $i \in \{1, \ldots, n\}$, we have* $\Pr\left[i = \arg\max_{j=1}^{n} x_j\right] \leq \varepsilon$, *where each x_i is a random variable drawn from D_i.*

In particular, we will show that our algorithm results in an improved bound (relative to the best single-threshold algorithm) when the set of distributions satisfies a no ε-superstars assumption for a sufficiently small constant ε. We will sometimes drop the ε and simply refer to the "no-superstars assumption" when ε is clear from context.

The starting point for our algorithm is the analysis of Gilbert and Mosteller 1966, which shows that in the i.i.d. setting the optimal (multi-threshold) algorithm succeeds with probability 0.5801 as n grows large. At an intuitive level, we would like to establish that a prophet secretary instance behaves similarly to an i.i.d. instance, where each of the distinct distributions is replaced by an "average" of all the distributions. However, this is not quite right due to correlations between values. For example, once the process reaches the last few distributions, the algorithm may have a lot of information about their likely outcomes relative to an i.i.d. instance, because knowing which distributions are left could be very informative.

To dampen this correlation, we will instead consider groups of qn consecutive observations for some small constant q. The maximum of each collection of qn distributions will, because of concentration from sampling, be distributed very similarly to the maximum of a suitable average of all the distributions, and there is negligible correlation between the $1/q$ collections. It is here where we make use of the no-superstars assumption. We can therefore model our best-choice prophet secretary instance as a (nearly) i.i.d. instance with $1/q$ observations, and design an algorithm based on the i.i.d. variation of the problem. This ultimately leads to an algorithm for best-choice prophet secretary that succeeds with probability as close as desired to the worst-case guarantee of the best i.i.d. algorithm.

Theorem 8 *Let Alg_τ be any threshold-based algorithm that selects the maximum with probability at least α when values are i.i.d. Then for any $\gamma \in (0, 1)$, there is an algorithm for the best-choice prophet secretary problem that selects the maximum with probability at least $(\alpha - 13\gamma)$, whenever the distributions satisfy the no ε-superstars assumption with $\varepsilon = \frac{\gamma^{10}}{24 \log(\frac{2}{\gamma^2})}$. In particular, for small enough ε, we can take $\alpha \approx 0.5801$.*

While Theorem 8 requires a no-superstars assumption, we can use it to show that for general input distributions, the single-threshold algorithm is not tight, under the additional assumption that we observe not just the value but also which distribution the value arises from in each observation.

Theorem 9 *There exists an algorithm for the best-choice prophet secretary problem that chooses the maximum value with probability at least $\max_\lambda \sum_{k=1}^{\infty} \left(\frac{1}{k} \frac{\lambda^k e^{-\lambda}}{k!}\right) + \varepsilon_0$, where ε_0 is a positive constant, when we observe not just the value but also the distribution from which each value arises.*

We give the formal details of our algorithm and analyze its success probability in Appendix C. The main technical difficulty in the analysis is establishing the necessary concentration bounds, which require some care because we are sampling without replacement and do not have a good uniform bound on the contribution of any single value. We defer the proof details of these concentration inequalities to Appendix E.1.

Acknowledgements

The First author is supported in part by NSF grants CCF-1320231 and CNS-1228598.

The second author is supported in part by NSF CAREER award CCF-1053605, NSF BIGDATA grant IIS-1546108, NSF AF:Medium grant CCF-1161365, DARPA GRAPHS/AFOSR grant FA9550-12-1-0423, and another DARPA SIMPLEX grant. Part of this work was done while visiting Microsoft Research New England.

The fourth author is supported in part by NSF grants CCF-1563710, CCF-1535795, CCF-1320231, and CNS-1228598. Part of this work was done while visiting Microsoft Research New England.

References

Melika Abolhassani, Soheil Ehsani, Hossein Esfandiari, MohammadTaghi HajiAghayi, Robert Kleinberg, and Brendan Lucier. Beating 1-1/e for ordered prophets. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 61–71. ACM, 2017.

Saeed Alaei. Bayesian combinatorial auctions: Expanding single buyer mechanisms to many buyers. *SIAM Journal on Computing*, 43(2):930–972, 2014.

Saeed Alaei, MohammadTaghi Hajiaghayi, and Vahid Liaghat. Online prophet-inequality matching with applications to ad allocation. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 18–35. ACM, 2012.

Moshe Babaioff, Nicole Immorlica, and Robert Kleinberg. Matroids, secretary problems, and online mechanisms. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 434–443. Society for Industrial and Applied Mathematics, 2007.

Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 311–320, 2010.

José Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Posted price mechanisms for a random stream of customers. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 169–186. ACM, 2017.

Eugene B Dynkin. The optimum choice of the instant for stopping a markov process. In *Soviet Math. Dokl*, volume 4, 1963.

Soheil Ehsani, MohammadTaghi Hajiaghayi, Thomas Kesselheim, and Sahil Singla. Prophet secretary for combinatorial auctions and matroids. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 700–714, 2018.

Hossein Esfandiari, MohammadTaghi Hajiaghayi, Vahid Liaghat, and Morteza Monemizadeh. Prophet secretary. In *To appear in SIAM Journal on Discrete Mathematics*. 2017.

Moran Feldman, Ola Svensson, and Rico Zenklusen. A simple O (log log (rank))-competitive algorithm for the matroid secretary problem. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1189–1201, 2015.

Thomas S. Ferguson. Who solved the secretary problem? *Statist. Sci.*, 4(3):282–289, 08 1989. doi: 10.1214/ss/1177012493.

P. R. Freeman. The secretary problem and its extensions: A review. *International Statistical Review*, 51 (2):189–206, 1983.

John P. Gilbert and Frederick Mosteller. Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.*, 61:35–73, 1966.

Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 982–991, 2008.

Guru Prashanth Guruganesh and Sahil Singla. Online matroid intersection: Beating half for random arrival. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 241–253. Springer, 2017.

S.M. Gusein-Zade. The problem of choice and the optimal stopping rule for a sequence of independent trials. *Theory of Probility and its Applications*, 11: 472–476, 1966.

Mohammad Taghi Hajiaghayi, Robert Kleinberg, and David C Parkes. Adaptive limited-supply online auctions. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 71–80. ACM, 2004.

Mohammad Taghi Hajiaghayi, Robert Kleinberg, and Tuomas Sandholm. Automated online mechanism design and prophet inequalities. In *AAAI*, volume 7, pages 58–65, 2007.

Chinmay Karande, Aranyak Mehta, and Pushkar Tripathi. Online bipartite matching with unknown distributions. In *Proceedings of the forty-third annual*

ACM symposium on Theory of computing, pages 587–596. ACM, 2011.

DP Kennedy. Prophet-type inequalities for multi-choice optimal stopping. *Stochastic Processes and their Applications*, 24(1):77–88, 1987.

DP Kennedy et al. Optimal stopping of independent random variables and maximizing prophets. *The Annals of Probability*, 13(2):566–571, 1985.

Robert P Kertz. Comparison of optimal value and constrained maxima expectations for independent random variables. *Advances in applied probability*, 18(02):311–340, 1986.

Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In *European Symposium on Algorithms*, pages 589–600. Springer, 2013.

Robert Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 630–631. Society for Industrial and Applied Mathematics, 2005.

Nitish Korula and Martin Pál. Algorithms for secretary problems on graphs and hypergraphs. In *International Colloquium on Automata, Languages and Programming*, pages 508–520. Springer, 2009.

Ulrich Krengel and Louis Sucheston. Semiamarts and finite values. *Bull. Am. Math. Soc*, 1977.

Ulrich Krengel and Louis Sucheston. On semiamarts, amarts, and processes with finite value. *Advances in Prob*, 4:197–266, 1978.

Oded Lachish. O (log log rank) competitive ratio for the matroid secretary problem. In *Proceedings of the Fifty-Fifth Annual IEEE Symposium on Foundations of Computer Science*, pages 326–335, 2014.

Lucien Le Cam. An approximation theorem for the poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197, 1960.

Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing lps. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 597–606. ACM, 2011.

Joseph D Petruccelli. Best-choice problems involving uncertainty of selection and recall of observations. *Journal of Applied Probability*, 18(2):415–425, 1981.

Zdzisław Porosiński. The full-information best choice problem with a random number of observations. *Stochastic Processes and their applications*, 24(2): 293–307, 1987.

J Michael Steele. Le cam's inequality and poisson approximations. *The American Mathematical Monthly*, 101(1):48–54, 1994.

Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

Qiqi Yan. Mechanism design via correlation gap. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 710–719. Society for Industrial and Applied Mathematics, 2011.