A Conclusion and Open Problems

We have provided several new results for the best-choice problem in the prophet inequality and prophet secretary settings, where the goal is to maximize the probability of selecting the largest value from a sequence of values drawn independently from known distributions. Many of our proofs involve Poissonization-style arguments, where we approximate the number of values above a threshold with a Poisson random variable. This approach was particularly useful for generalizing results from the i.i.d. setting to the different setting of arbitrary distributions in a random order. We believe this approach may be useful for other related problems.

Our main open problems relate to our most technical result, namely that, under the no superstars assumption, we can use an algorithm with multiple thresholds to select the maximum with probability approximately 0.5801 in the setting with arbitrary distributions in a random order. It is open to determine what probability can be achieved for arbitrary distributions in a random order without the no superstars assumption. A related open question would be to simplify our proof; it would be interesting to know if there is a more straightforward argument, and such an argument might more readily lead to results without the no superstars assumption. Indeed, we conjecture the following: that for any n, the worst-case instance of the best-choice prophet secretary problem is an i.i.d. instance, so in particular the worst-case success probability matches that of the i.i.d. best-choice problem.

Another open problem is to consider "best-case" orderings, where the player trying to select the maximum is allowed to choose the order of the distributions for observation. Does the ability to choose the ordering provide an advantage over random order, in the worst case? Even beyond worst-case instances, there is a computational problem of finding the best ordering. Can the best ordering for an arbitrary problem instance be found in polynomial time?

We extended our results to the problem of selecting one of the top k values. More generally, one could consider the problem of maximizing other functions of the rank of the value selected, such as minimizing the expected rank. One could also study variants in which multiple values can be selected, subject to a downward-closed constraint, and the goal is to maximize a function of the set of ranks of the selected values. For example, how should one select values subject to a matroid constraint, so as to maximize the probability that the largest value is among the values selected?

B Proof of Theorem 2

Next we prove Theorem 2.

Consider the following example. There are n random variables $x_1, \dots, x_n$ from distributions $D_1, \dots, D_n$ as follows: for any $i \in \{1, \cdots, n\}$, x_i is i with probability $q_i = \frac{1}{i}$ and 0 otherwise.[6] Note that the nonzero random variable with the largest index is the maximum. Hence the probability of x_i being the maximum is independent of the x_j values with $j < i$. Moreover, x_1 is always 1 and hence the maximum is never 0. We let p_i be the probability that x_i is the maximum. The distributions will arrive in index order.

We claim that $p_1 = \cdots = p_n = \frac{1}{n}$. We show this by strong induction.[7] The base case holds for $i = n$ where $p_n = q_n = \frac{1}{n}$. Assuming $p_{i+1} = \cdots = p_n = \frac{1}{n}$, we have

$$\begin{aligned} p_i &= q_i \cdot \Pr\left[x_{i+1} = \cdots = x_n = 0\right] \\ &= q_i \cdot \Pr\left[\text{maximum is not in } \{x_{i+1}, \dots, x_n\}\right] \\ &= q_i\Big(1 - \sum_{j=i+1}^{n} p_j\Big) = \frac{1}{i}\frac{i}{n} = \frac{1}{n}. \end{aligned}$$

Hence we have $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$. Also, we have $\Pr\left[x_i = \max_{j=1}^{n} x_j | x_i \neq 0\right] = \frac{i}{n}$.

Let Alg be the best online algorithm and let $\mathsf{Alg}_{i+1\to n}$ be the probability that Alg picks the maximum assuming that it rejects $x_1, \dots, x_i$. Notice that if Alg picks a nonzero number x_j from $\{x_{i+1}, \dots, x_n\}$, it is larger than all

[6] One can make this atomless by assuming that x_i is drown uniformly at random from $[i, i+\epsilon]$ with probability $q_i = \frac{1}{i}$ and 0 otherwise.

[7] This follows similar reasoning to the analysis of reservoir sampling.

numbers in $\{x_1, \dots, x_i\}$. Hence $\mathsf{Alg}_{i+1\to n}$ is independent of $x_1, \dots, x_i$. Notice that if Alg rejects x_{i+1} it picks the maximum with probability $\mathsf{Alg}_{i+2\to n}$. Hence $\mathsf{Alg}_{i+1\to n} \geq \mathsf{Alg}_{i+2\to n}$, which means $\mathsf{Alg}_{i+1\to n}$ is decreasing in i.

Indeed, if $x_i \neq 0$, when $\Pr\left[x_i = \max_{j=1}^n x_j | x_i \neq 0\right] \geq \mathsf{Alg}_{i+1\to n}$, Alg picks x_i and stops. Otherwise, Alg rejects x_i and continues. Also remember that $\Pr\left[x_i = \max_{j=1}^n x_j | x_i \neq 0\right]$ is increasing in i and $\mathsf{Alg}_{i+1\to n}$ is decreasing in i. Therefore, there exists an index i such that for all $j < i$, Alg rejects x_i and accepts the first nonzero x_j with $j \geq i$. Therefore, Alg picks the maximum with probability

$$
\begin{aligned}
&\sum_{j=i}^{n} \Pr\left[x_i = \cdots = x_{j-1} = 0\right] \Pr\left[x_j = \max_{j'=1}^{n} x_{j'}\right] \\
&= \frac{1}{n} + \sum_{j=i+1}^{n} \Pi_{k=i}^{j-1}(1 - 1/k)\frac{1}{n} \\
&= \frac{1}{n} + \sum_{j=i+1}^{n} \frac{i-1}{j-1}\frac{1}{n} \\
&= \frac{1}{n} + \frac{i-1}{n}\sum_{j=i+1}^{n} \frac{1}{j-1} \\
&\leq \frac{1}{n} + \frac{i-1}{n}\left(\ln(\frac{n+1}{i}) + \frac{1}{i}\right) \\
&\leq \frac{2}{n} + \frac{i-1}{n}\ln(\frac{n+1}{i}) \\
&\leq \frac{2}{n} + \frac{i}{n+1}\ln(\frac{n+1}{i}) \\
&\leq \frac{2}{n} + \alpha\ln(1/\alpha) \qquad \text{for } \alpha = \frac{i}{n+1} \in [0,1].
\end{aligned}
$$

Note that $\alpha\ln(1/\alpha)$ maximizes at $\alpha = \frac{1}{e}$. Thus, Alg picks the maximum with probability at most $\frac{1}{e} + \frac{2}{n}$.

C An Algorithm for Best-Choice Prophet Secretary

Before we describe our algorithm for the best-choice prophet secretary problem, we must first provide some definitions and fix some parameters. Throughout this subsection, for an arbitrary $\gamma \in (0,1)$ we set $\lambda_0 = \gamma$, $\rho = \gamma^3$, $q = \frac{\gamma^2}{2}$, and $\delta = \frac{\gamma^6}{4}$. Notice that we have $\frac{\gamma\lambda_0}{2\rho} = \frac{\gamma^2}{2\gamma^3} = \frac{1}{2\gamma} \geq \frac{\gamma^2}{2} = q$. We will then set $\varepsilon = \frac{\gamma^2 q^2 \rho\lambda_0}{2\log\frac{2}{\delta}} = \frac{\gamma^{10}}{8\log(\frac{8}{\gamma^6})} = \frac{\gamma^{10}}{24\log(\frac{2}{\gamma^2})}$; this will be the value ε we require in the no-superstars assumption. We note that we have not aimed to optimize these parameters.

Set $c = \frac{1-\lambda_0}{\rho}$. We let $t_0, \dots, t_c$ be the (unique) sequence of thresholds such that, for each $\zeta \in \{0, \dots, c\}$, we have $\Pr\left[\max_{i=1}^n(x_i) \leq t_\zeta\right] = \lambda_0 + \zeta\rho$. That is, the probability that $\max_{i=1}^n(x_i)$ falls between any two consecutive thresholds is ρ, and the probability that it falls below t_0 is λ_0.

The next definition captures our desire to combine multiple distributions D_i into a single collection, and study the maximum of the values drawn from that collection of distributions.

Definition 10 (Collection Distribution) *Let $S \subseteq \{D_1, \dots, D_n\}$ be an arbitrary set. We define the collection distribution D_S using the following procedure: D_S draws x_i from distribution D_i for each $D_i \in S$, then returns $\max_{D_i\in S} x_i$. We use x_S to indicate an outcome of D_S.*

The following lemma provides a concentration result for the distribution D_S, when S is a set of size qn chosen uniformly at random without replacement from $D_1, \dots, D_n$. Intuitively, this says that if we decompose a random order sequence of $D_1, \dots, D_n$ into $\frac{1}{q}$ subsequences, each of size qn, these subsequences behave similarly to an i.i.d. distribution. We use this to prove our main result. We defer the proof to Section E.1.

Algorithm 1:

Parameters: Thresholds $\tau_1, \ldots, \tau_{1/q}$

Input: Iteratively receive values $\tilde{x}_{S_\eta}$, for $\eta \in \{1, \ldots, \frac{1}{q}\}$.

1: With probability 4γ, do not pick $\tilde{x}_{S_\eta}$ and move to the next number.
2: Set t_0 such that $\Pr[\max_{i=1}^{n}(x_i) \le t_0] = \lambda_0$.
3: **if** $\tilde{x}_{S_\eta} \le t_0$ **then**
4: Do not pick $\tilde{x}_{S_\eta}$ and move to the next number.
5: **if** $\tilde{x}_{S_\eta} \le \tau_\eta$ **then**
6: Do not pick $\tilde{x}_{S_\eta}$ and move to the next number.
7: **else**
8: Pick $\tilde{x}_{S_\eta}$.

Lemma 11 *Let* S *be a set of size* qn, *chosen uniformly at random without replacement from* $D_1, \ldots, D_n$. *With probability* $1 - \frac{\gamma^3}{2}$ *for all* $\zeta \in \{0, \ldots, c-1\}$ *we have*

$$(1-3\gamma)q\sum_{i=1}^{n} p_i^\zeta \le \Pr[t_\zeta \le x_S < t_{\zeta+1}] \le (1+\gamma)q\sum_{i=1}^{n} p_i^\zeta,$$

where $p_i^\zeta = \Pr[t_\zeta \le x_i < t_{\zeta+1}]$, *assuming the no* ε*-superstars assumption with* $\varepsilon = \frac{\gamma^{10}}{24\log(\frac{2}{\gamma^2})}$.

We use the following definitions in our proof of Theorem 18.

Definition 12 *For a given number* $x \ge t_0$, *we write* $\tilde{x} = \max\{t_\zeta : t_\zeta \le x\}$. *That is,* $\tilde{x}$ *is* x *rounded down to the nearest* t_ζ. *Similarly, for a distribution* D *we use* $\tilde{D}$ *to represent the distribution that draws* x *from* D *and then returns* $\tilde{x}$.

Definition 13 *We define a distribution* $D_{\min}$ *as follows: for any* $\zeta \in \{0, \ldots, c-1\}$, $D_{\min}$ *returns* t_ζ *with probability* $(1-3\gamma)q\sum_{i=1}^{n} p_i^\zeta$, *and otherwise* $D_{\min}$ *returns* 0.

Definition 14 *For* $\eta \in \{1, \ldots, \frac{1}{q}\}$, *let* S_η *be the set of distributions* $D_{\pi_{(\eta-1)qn+1}}, \ldots D_{\pi_{(\eta)qn}}$. *Let* $\hat{D}_{S_\eta}$ *be a distribution that returns* $\tilde{x}_{S_\eta}$ *with probability* $1-4\gamma$ *and returns* 0 *otherwise. We use* $\hat{x}_{S_\eta}$ *to indicate an outcome of* $\hat{D}_{S_\eta}$.

We now present results for two algorithms, Algorithm 1 and Algorithm 2, whose pseudocode is listed in the text. These algorithms take, as parameters, a sequence of thresholds defining an arbitrary threshold-based algorithm for the i.i.d. setting with $1/q$ observations. Algorithm 1 provides an intermediary result. In particular, Algorithm 1 is meant to work with the values $\tilde{x}_{S_\eta}$, which recall are "rounded down" values drawn from the collection distributions. This algorithm is used to bound the success rate if we used the $1/q$ collection distributions to generate our input instead of the actual observations. We then show that Algorithm 2, which works with the real observations, performs nearly as well as Algorithm 1.

We first show that Algorithm 1 can simulate an arbitrary i.i.d. algorithm with minimal loss, under a no-superstars assumption.

Lemma 15 *Let* Alg_τ *be any threshold-based algorithm that selects the maximum with probability at least* α *for* $1/q$ *instances of* $D_{\min}$, *with thresholds* $\tau_1, \ldots, \tau_{1/q}$. *For any arbitrary* $\gamma \in (0,1)$, *Algorithm 1 selects the maximum with probability at least* $(\alpha - 10\gamma)$ *for* $\tilde{D}_{S_1}, \ldots, \tilde{D}_{S_{1/q}}$, *assuming the no* ε*-superstars assumption with* $\varepsilon = \frac{\gamma^{10}}{24\log(\frac{2}{\gamma^2})}$.

Proof : First of all notice that the probability that the maximum is less than t_0 is $\lambda_0 = \gamma$. We assume that any number less than t_0 is 0 and we do not pick it. We miss the maximum with probability γ due to this assumption. Algorithm 1 handles this assumption by the condition in line 3.

Algorithm 2:

Parameters: Thresholds $\tau_1, \dots, \tau_{1/q}$
Input: Iteratively receive values x_{π_i}, for $i \in \{1, \dots, n\}$.
1: With probability 4γ, do not pick x_{π_i} and move to the next number.
2: Set t_0 such that $\Pr[\max_{i=1}^{n}(x_i) \le t_0] = \lambda_0$.
3: **if** $\tilde{x}_{\pi_i} \le t_0$ **then**
4: Do not pick x_{π_i} and move to the next number.
5: **if** $\tilde{x}_{\pi_i} \le \tau_{\lceil qi \rceil}$ **then**
6: Do not pick x_{π_i} and move to the next number.
7: **else**
8: Pick x_{π_i}.

By Lemma 11 with probability $1 - \frac{\gamma^3}{2}$ for all $\zeta \in \{0, \dots, c\}$ we have

$$\Pr\left[\hat{x}_{S_\eta} = t_\zeta\right] \le \Pr\left[x_{\min} = t_\zeta\right] \le \Pr\left[\tilde{x}_{S_\eta} = t_\zeta\right], \tag{3}$$

where the first inequality follows from $(1-4\gamma)(1+\gamma) \le 1-3\gamma$ (where $1+\gamma$ and $1-3\gamma$ are coming from Lemma 10 and $1-4\gamma$ is coming from the definition of $\hat{x}_{S_\eta}$ i.e. Definition 13). By the union bound this holds for all $\eta \in \{1, \dots, \frac{1}{q}\}$ and all $\zeta \in \{0, \dots, c\}$ with probability at least $1 - \frac{1}{q}\frac{\gamma^3}{2} = 1-\gamma$. In the rest of the proof we assume that Inequality 3 holds for all $\eta \in \{1, \dots, \frac{1}{q}\}$ and all $\zeta \in \{0, \dots, c\}$.

We define ϕ_η to be the probability that Alg_τ reaches the η-th number when running on $\frac{1}{q}$ instances of $D_{\min}$. Similarly, we define $\tilde{\phi}_\eta$ to be the probability Algorithm 1 reaches the η-th number when running on $\tilde{D}_{S_1}, \dots, \tilde{D}_{S_{1/q}}$. We also define σ_η to be the probability that Algorithm Alg_τ, conditioned on reaching the η-th number, accepts the η-th number when running on $\frac{1}{q}$ instances of $D_{\min}$ and succeeds. Similarly, we define $\tilde{\sigma}_\eta$ to be the probability Algorithm 1, conditioned on reaching the η-th number, accepts the η-th number when running on $\tilde{D}_{S_1}, \dots, \tilde{D}_{S_{1/q}}$ and succeeds. We refer to this notion as the probability of success at η. Notice that the probability that Alg_τ and Algorithm 1 succeed are $\sum_{\eta=1}^{1/q} \phi_\eta \sigma_\eta$ and $\sum_{\eta=1}^{1/q} \tilde{\phi}_\eta \tilde{\sigma}_\eta$ respectively.

In fact, running Algorithm 1 on $\tilde{D}_{S_1}, \dots, \tilde{D}_{S_{1/q}}$ is equivalent to running Lines 3 to 8 on $\hat{D}_{S_1}, \dots, \hat{D}_{S_{1/q}}$. Hence by inequality 3 we have

$$\sum_{\eta=1}^{1/q} \tilde{\phi}_\eta \tilde{\sigma}_\eta \ge \sum_{\eta=1}^{1/q} \phi_\eta \tilde{\sigma}_\eta.$$

Now, let $\eta \in \{1, \dots, \frac{1}{q}\}$ be an arbitrary index. Assume for all $\eta' \in \{1, \dots, \frac{1}{q}\} \setminus \{\eta\}$ we replace distributions $\tilde{D}_{S_{\eta'}}$ with $D_{\min}$. By Inequality 3 this increases the the probability of success at η by at most a factor $\frac{1}{1-4\gamma}$. Next, if we replace $\tilde{D}_{S_\eta}$ with $D_{\min}$ the probability of success at η decreases and becomes $(1-4\gamma)\sigma_\eta$. Thus, we have $\frac{1}{1-4\gamma}\tilde{\sigma}_\eta \ge (1-4\gamma)\sigma_\eta$, which implies $\tilde{\sigma}_\eta \ge (1-4\gamma)^2\sigma_\eta \ge (1-8\gamma)\sigma_\eta$. Therefore we have

$$\sum_{\eta=1}^{1/q} \phi_\eta \tilde{\sigma}_\eta \ge (1-8\gamma) \sum_{\eta=1}^{1/q} \phi_\eta \sigma_\eta \ge (1-8\gamma)\alpha \ge \alpha - 8\gamma.$$

Remember that as we mentioned in the beginning, Algorithm 1 misses the maximum with probability γ due to the condition in line 3, and it loses another γ probability by assuming that Inequality 3 holds for all $\eta \in \{1, \frac{1}{q}\}$ and all $\zeta \in \{0, \dots, c\}$. Hence the probability of selecting the maximum drops to $\alpha - 10\gamma$. □

We now want to prove that Algorithm 2 can likewise simulate an arbitrary i.i.d. algorithm with minimal loss, by comparing to the performance of Algorithm 1. Recall that Algorithm 2 attempts to simulate Algorithm 1 by applying threshold τ_η to each of the qn values in collection η. There are two ways that this simulation might fail. First, it might be that two values in collection η are above threshold τ_η, and Algorithm 2 chooses the smaller one. Second, it could be that the maximum value from two different collections both round to the same value

$\tilde{x}$, and Algorithm 1 chooses the smaller one; this is fine for Algorithm 1, since it cares only about the rounded values, but leads to failure for Algorithm 2.

The following two concentration results handle these two modes of failure. Lemma 16 shows that it is unlikely that two or more values in any given collection lie above the corresponding threshold. Lemma 17 shows that it is unlikely that the maximum value in two different collections round to the same t_ζ. We defer the proofs to Section E.1.

Lemma 16 *Consider arbitrary numbers* $\lambda_0, \gamma, \delta, q \in (0,1)$*,* $\rho \in (0, 1-\lambda_0)$*. Set* $\varepsilon = \frac{\gamma^2 q^2 \rho \lambda_0}{2 \log \frac{2}{\delta}}$*. Let* S *be a set of size* qn*, chosen uniformly at random without replacement from* $D_1, \dots, D_n$*. Let* τ^0 *be such that* $\Pr\left[\max_{i=1}^n(x_i) \le \tau^0\right] = 1 - \rho$*. Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \le x_i$ *and* 0 *otherwise. Let* $p'_i = \Pr[y_i = 1]$*. Assuming the no* ε*-superstars assumption, with probability* $1-\delta$ *we have*

$$\Pr[\exists_{i \in S} y_i = 1] \le \frac{2q}{\lambda_0}$$

and

$$\Pr\left[\sum_{i \in S} y_i \ge 2\right] \le \frac{4q^2}{\lambda_0^2}.$$

Lemma 17 *Consider arbitrary numbers* $\rho, \lambda_0 \in (0,1)$ *and* $\lambda \in [0, 1-(\lambda_0+\rho)]$*. Let* τ^0 *and* τ^1 *be such that* $\Pr\left[\max_{i=1}^n(x_i) \le \tau^0\right] = 1 - (\lambda+\rho)$ *and* $\Pr\left[\max_{i=1}^n(x_i) \le \tau^1\right] = 1 - \lambda$*. Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \le x_i \le \tau^1$ *and* 0 *otherwise. We have*

$$\Pr\left[\sum_{i=1}^n y_i \ge 2\right] \le \frac{\rho^2}{\lambda_0^2}.$$

These lemmas in hand, we are now ready to bound the success probability of Algorithm 2. This is Theorem 18, which was a restatement of our main result for the best-choice prophet secretary problem under a no-superstars assumption, Theorem 8.

Theorem 18 *Let* Alg_τ *be a threshold based algorithm that selects the maximum with probability at least* α *for* $1/q$ *instances of* $D_{\min}$*, with thresholds* $\tau_1, \dots, \tau_{1/q}$*. For any arbitrary* $\gamma \in (0,1)$*, Algorithm 2 selects the maximum with probability at least* $(\alpha - 13\gamma)$ *for* $D_{\pi_1}, \dots, D_{\pi_n}$*, assuming the no* ε*-superstars assumption with* $\varepsilon = \frac{\gamma^{10}}{24 \log(\frac{2}{\gamma^2})}$*.*

Proof : There are two basic differences between Algorithm 1 and Algorithm 2. First, for each of the sets of qn consecutive numbers S_η, Algorithm 1 has the privilege to observe the maximum number in the set at once, while Algorithm 2 sees the numbers in the set one by one. Second, the input numbers in Algorithm 1 are all rounded to t_ζ's, but this is not true for the input of Algorithm 2. Therefore, there are two cases where Algorithm 1 selects the maximum of the $\tilde{x}_{S_\eta}$ but Algorithm 2 does not choose the maximum of the x_{π_i}.

- Algorithm 1 picks $\tilde{x}_{S_\eta}$. There are two numbers $\tau_\eta < x_i < x_{i'}$ with $i, i' \in S_\eta$, and Algorithm 2 picks x_i.
- Algorithm 1 picks $\tilde{x}_{S_\eta}$. But there is another η' such that $\tilde{x}_{S_\eta} = \tilde{x}_{S_{\eta'}} = t_\zeta$ but $x_{S_\eta} < x_{S_{\eta'}}$.

We show that first case happens with probability at most 2γ and the second case happens with probability at most γ. This together with Lemma 15 proves the theorem. Notice that the probability of the first case is at most

$$\begin{aligned}\Pr\left[\exists_{i' \in S_\eta \setminus \{i\}} x_{i'} \ge \max(\tau_\eta, t_0)\right] &\le \Pr\left[\exists_{i' \in S_\eta} x_{i'} \ge \max(\tau_\eta, t_0)\right] \\ &\le \frac{2q}{\lambda_0} = \gamma, \qquad \text{By Lemma 16}\end{aligned}$$

where Lemma 16 holds with probability $1 - \delta \ge 1 - \gamma$. Hence the first case happens with probability at most $\gamma + \gamma = 2\gamma$.

Notice that in the second case for some ζ there are at least two numbers x_i (corresponds to η) and $x_{i'}$ (corresponds to η') such that $t_\zeta \leq x_i \leq x_{i'} \leq t_{\zeta+1}$. By Lemma 17, for a particular ζ this happens with probability at most $\frac{\rho^2}{\lambda_0^2}$. By the union bound over all choices of ζ, the second case happens with probability at most $c\frac{\rho^2}{\lambda_0^2} \leq \frac{\rho}{\lambda_0^2} = \frac{\gamma^3}{\gamma^2} = \gamma$.
□

Now we are ready to prove Theorem 9, which is an unconditional improvement that holds even without the no-superstars assumption.

Proof of Theorem 9: By Theorem 8, there is a positive constant $\varepsilon > 0$ such that the statement of Theorem 9 holds whenever the distributions satisfy the no ε-superstars assumption. We will therefore assume that there exists a distribution in the input that violates the no ε-superstars assumption for this positive constant ε. That is, $\Pr\left[i = \arg\max_{j=1}^n x_j\right] \geq \varepsilon$ for some i. Without loss of generality we assume that this distribution is D_1. Let τ be the threshold selected by the algorithm in Theorem 4. Recall that Theorem 4 shows that, for any arbitrary $\varepsilon' > 0$, there exists a single threshold algorithm that chooses the maximum value with probability at least $\max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right) - \varepsilon'$, for the best-choice prophet secretary problem. For the purpose of this theorem, we set $\varepsilon' = \frac{e^{-1.5}\varepsilon^2}{32}$. We will consider two cases. In the first case we have $\Pr\left[x_1 < \tau \text{ and } 1 = \arg\max_{j=1}^n x_j\right] \geq \frac{\varepsilon}{2}$. In the second case we have $\Pr\left[x_1 \geq \tau\right] \geq \frac{\varepsilon}{2}$. Note that we must be in one of these cases, since

$$\Pr\left[x_1 < \tau \text{ and } 1 = \arg\max_{j=1}^n x_j\right] + \Pr\left[x_1 \geq \tau\right] \geq \Pr\left[1 = \arg\max_{j=1}^n x_j\right] \geq \varepsilon.$$

Case 1. In this case we apply the single threshold algorithm of Theorem 4, with a slight modification: if D_1 is one of the last $\frac{\varepsilon n}{2}$ items, and we reach it, we stop and accept it regardless of its value. Note that the probability that D_1 appears in one of the last $\frac{\varepsilon n}{2}$ positions, and at the same time the maximum appears after D_1 (and hence also somewhere in the last $\frac{\varepsilon n}{2}$ positions), is at most $\frac{\varepsilon}{2} \times \frac{\varepsilon}{2} \times \frac{1}{2} = \frac{\varepsilon^2}{8}$. This is an upper bound on the loss of using this modification of the algorithm. On the other hand, the probability that D_1 appears as one of the last $\frac{\varepsilon n}{2}$ items, is the maximum item, and is below the threshold τ (which also means no item is above the threshold) is at least $\Pr\left[x_1 < \tau \text{ and } 1 = \arg\max_{j=1}^n x_j\right] \times \frac{\varepsilon}{2} \geq \frac{\varepsilon^2}{4}$. This is a lower bound on the expected gain of using this modification to the algorithm. Therefore in this case we improve Theorem 4 by at least $\frac{\varepsilon^2}{4} - \frac{\varepsilon^2}{8} = \frac{\varepsilon^2}{8}$.

Case 2. In this case we show that the analysis of Theorem 4 in not tight and hence we provide a better bound for the algorithm with threshold τ. To prove this, we show a constant gap in Inequality 2, which directly translates to a constant improvement on the probability of success of the algorithm. Specifically, we consider the case where D_1 is the only item above the threshold, but more than one of its corresponding dummy distribution is above the threshold (i.e., $\mathcal{K}' \geq 2$). In this situation, the algorithm certainly selects the maximum; however, in the analysis, we assumed that of the $\mathcal{K}'$ values above the threshold from the dummy distributions, the algorithm would only choose the maximum with probability $\frac{1}{\mathcal{K}'} \leq \frac{1}{2}$ due to the ordering of items. Recall that $\Pr\left[\max_{i=1}^n(x_i) \leq \tau\right] = e^{-\lambda} > e^{-1.5}$ and hence $\Pr\left[\max_{i=2}^n x_i \leq \tau\right] > e^{-1.5}$. Moreover, note that $\frac{\Pr[x_1 \geq \tau]}{2}$ is a lower bound on the probability that we see at least one item above the threshold in half of the dummy distribution corresponding to D_1 and hence with probability at least $\left(\frac{\Pr[x_1 \geq \tau]}{2}\right)^2$ we see at least one item above the threshold in the first half of the distributions and at least one in the second half. Thus, we have

$$\Pr\left[\mathcal{K}' \geq 2 \text{ and } x_1 \geq \tau \text{ and } \forall_{i\in\{2,\ldots,n\}} x_i < \tau\right] \geq \left(\frac{\Pr\left[x_1 \geq \tau\right]}{2}\right)^2 \times \Pr\left[\forall_{i\in\{2,\ldots,n\}} x_i < \tau\right] \geq \frac{e^{-1.5}\varepsilon^2}{8},$$

Therefore, in an event that occurs with probability at least $\frac{e^{-1.5}\varepsilon^2}{8}$, we can improve our bound from something at most $\frac{1}{2}$ to 1. This leads to a gap of $\frac{e^{-1.5}\varepsilon^2}{16}$ in Inequality 2, and hence a corresponding improvement to Theorem 4.

Thus, in either case, we obtain an improvement of $\epsilon_0 = \frac{\varepsilon^2}{16e^{1.5}}$ to the bound in Theorem 4, which says we select the maximum value with probability at least $\max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right) - \varepsilon' + \frac{\varepsilon^2}{16e^{1.5}} = \max_\lambda \sum_{k=1}^{\infty}\left(\frac{1}{k}\frac{\lambda^k e^{-\lambda}}{k!}\right) + \frac{\varepsilon^2}{32e^{1.5}}$.
□

D Appendix: Omitted Proofs from Section 5

We present the proof of Theorem 5, which states that one can solve the top-k-choice prophet inequality problem with a failure rate that is exponentially decreasing in k. We restate the theorem below for completeness.

Theorem 19 *For any $k \geq 1$, there exists a single-threshold algorithm for the top-k-choice prophet inequality problem that succeeds with probability at least $1 - 2e^{-\gamma k}$, where $\gamma = (3 - \sqrt{5})/2$.*

Proof : We'll begin by showing a bound with a slightly worse constant in the exponent. We will then describe a way to optimize the constant at the end of the proof.

For a given constant t, let $X(t)$ be the random variable corresponding to the number of items i such that $x_i \geq t$. Choose τ so that $\mathrm{E}[X(\tau)] = k/2$.

The single threshold algorithm with threshold τ will succeed unless $X(\tau) = 0$ or $X(\tau) > k$. We note that $X(\tau)$ is the sum of n Bernoulli random variables, where variable i is 1 with probability $\Pr[x_i \geq t]$. By the additive form of the Chernoff bound, we have that

$$\Pr[X(\tau) = 0] = \Pr[X(\tau) \leq \mathrm{E}[X(\tau)] - k/2] < e^{-\mathsf{KL}(0||k/2n)\cdot n}$$

where $\mathsf{KL}(p||q)$ denotes the Kullback-Leibler (KL) divergence. Using the bound $\mathsf{KL}(p||q) \geq (p-q)^2/q$ for $p < q$, we have that

$$\Pr[X(\tau) = 0] < e^{-\mathsf{KL}(0||k/2n)\cdot n} < e^{n\cdot(k/2n)^2/(k/2n)} = e^{-k/4}.$$

Similarly, we have

$$\Pr[X(\tau) > k] = \Pr[X(\tau) > \mathrm{E}[X(\tau)] - k/2] < e^{-\mathsf{KL}(k/n||k/2n)\cdot n} < e^{n\cdot(k/2n)^2/(k/n)} = e^{-k/2}$$

where the second inequality uses the bound $\mathsf{KL}(p||q) \geq (p-q)^2/p$ for $p > q$. Taking a union bound over these two events completes the proof.

We note that if we choose a threshold τ so that $\mathrm{E}[X(\tau)] = \gamma k$ for $\gamma = (3 - \sqrt{5})/2$, we obtain a slightly better probability of success $1 - 2e^{-\gamma k}$ with the same argument. We have not sought to optimize the constant further. □

We next present the proof of Theorem 6, which shows that one cannot improve upon this exponential dependence on k, regardless of n and even for i.i.d. instances. We restate the theorem below.

Theorem 20 *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 $1 - e^{-c\cdot k}$.*

Proof : Take $n > k$ sufficiently large. Our problem instance is i.i.d., with distribution D as follows. With probability k/n, distribution D takes a value drawn uniformly from $[1, 2]$; with the remaining probability, the value is 0. We say that an observation is *successful* if it takes on a non-zero value. In order to describe our analysis more conveniently, we will think of the random process that generates our sequence of observations in the following alternative—but equivalent—way.

- We first draw n values uniformly from $[1, 2]$, say $v_1 < v_2 < \ldots < v_n$. We think of v_i as the value that x_i will take if x_i is non-zero. We write D_i for the distribution that takes on value v_i with probability k/n and 0 otherwise. We will think of value x_i as being drawn from distribution D_i.
- We choose a permutation π on $\{1, \cdots, n\}$; $\pi(i)$ is the position in the sequence that distribution D_i appears.
- We choose a number of successes Z_1 for the first $n/2$ observations, and correspondingly a number of successes Z_2 for the second $n/2$ observations. Both Z_1 and Z_2 are binomial random variables $\mathrm{Bin}(n/2, k/n)$ and are chosen accordingly.
- We choose permutations σ_1 on $\{1, \cdots, n/2\}$ and σ_2 on $\{n/2 + 1, \cdots, n\}$; σ_1 gives the order of the successful observations in the first $n/2$ observations, and similarly for σ_2, as described below.

More formally, we see observations in the order $x_{\pi(1)}, \ldots, x_{\pi(n)}$. For each $t \in \{1, \cdots, n/2\}$, $x_{\pi(t)} = v_{\pi(t)}$ if $\sigma_1(t) \leq Z_1$, and otherwise $x_{\pi(t)} = 0$. Similarly, for each $t \in \{n/2+1, \cdots, n\}$, $x_{\pi(t)} = v_{\pi(t)}$ if $\sigma_2(t) \leq Z_2$, and otherwise $x_{\pi(t)} = 0$. This process generates a distribution over value sequences that is identical to the distribution of value sequences in our i.i.d. top-k-choice problem.

We now consider the following events. Event A is that $Z_1 = k$; that is, the first half has k non-zero values. Event B is that, for each t_1, t_2 satisfying $t_1 \leq n/2$, $t_2 > n/2$, $\sigma_1(t_1) \leq k$, and $\sigma_2(t_2) \leq k$, we have that $\pi(t_1) \leq \pi(t_2)$. That is, event B is that the first k non-zero values in the first half of the observations (as determined by σ_1) will be less than the first k non-zero values in the second half (as determined by σ_2). Note that, from the way we have defined event B, it is independent of Z_1 and Z_2, as it depends only on π, σ_1, and σ_2. Because of this, events A and B are independent of each other (and independent of the value of Z_2).

We make the following claims. First, each of the events A and B happen with probability $e^{-\theta(k)}$. Second, conditioned on both A and B occurring, any algorithm must fail with probability at least $e^{-\theta(k)}$. The result follows immediately from these claims.

For event A, Z_1 is distributed as $\text{Bin}(n/2, k/n)$, and a simple calculation shows that it equals k with probability at least $e^{-c_1 k}$ for a suitable constant c_1 and large enough k. Indeed, the distribution is well approximated by a Poisson distribution, so the desired probability is approximately $e^{-k/2}(k/2)^k/k!$, which is $e^{-\theta(k)}$.

For event B, since π is a random ordering on the elements, the probability the first k values determined by σ_1 are all less than the first k values determined by σ_2 is just $\binom{2k}{k} \approx 2^{2k}/\sqrt{\pi k}$, which is $e^{-\theta(k)}$.

Now, for any algorithm, consider any realization of $\{v_1, \ldots, v_n\}$, π, σ_1, σ_2, and Z_1 for which events A and B both occur. Note that specifying Z_2 then specifies the entire process. Let us give the algorithm the additional power to decide, knowing $\{v_1, \ldots, v_n\}$, π, σ_1, σ_2, and Z_1 (but not Z_2), whether to have selected an element or not after the first $n/2$ observations. If the algorithm does not select an item, it will fail when $Z_2 = 0$, as then the k largest items have all appeared in the first half. If the algorithm does select an item, it will fail when $Z_2 \geq k$, as then the k largest items all appear in the second half. As Z_2 is distributed as $\text{Bin}(n/2, k/n)$, each of these possibilities for Z_2 occurs with probability $e^{-\theta(k)}$. Thus, if we condition on A and B both occurring, the algorithm fails with probability $e^{-\theta(k)}$ whether or not it chooses a value from among the first $n/2$ observations, and the result follows. $\square$

E Appendix: Omitted Proofs from Appendix C

E.1 Concentration Bounds

This section is dedicated to the proofs of Lemmas 11, 16, and 17. To begin, we require several prelimiary lemmata. The following lemma, for an arbitrary pair of thresholds $\tau^0 \leq \tau^1$, bounds the probability that at least one of the x_i's is within the range $[\tau^0, \tau^1]$.

Lemma 21 *Consider arbitrary numbers* $\rho \in (0,1)$ *and* $\lambda \in [0, 1-\rho)$. *Let* τ^0 *and* τ^1 *be such that* $\Pr\left[\max_{i=1}^n(x_i) \leq \tau^0\right] = 1-(\lambda+\rho)$ *and* $\Pr\left[\max_{i=1}^n(x_i) \leq \tau^1\right] = 1-\lambda$. *Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \leq x_i \leq \tau^1$ *and* 0 *otherwise. We have*

$$\rho \leq \Pr\left[\exists_{i \in \{1,\ldots,n\}} y_i = 1\right] \leq \frac{\rho}{1-\lambda}.$$

Proof : On one hand we have

$$\Pr\left[\exists_{i \in \{1,\ldots,n\}} y_i = 1\right] \geq \Pr\left[\tau^0 \leq \max_{i=1}^n(x_i) \leq \tau^1\right] = \rho.$$

On the other hand we have

$$\begin{aligned}
\lambda + \rho &= \Pr\left[\max_{i=1}^{n}(x_i) > \tau^0\right] \\
&= \Pr\left[\max_{i=1}^{n}(x_i) > \tau^1\right] + \Pr\left[\max_{i=1}^{n}(x_i) \le \tau^1\right] \times \Pr\left[\exists_{i\in\{1,\ldots,n\}}\tau^0 \le x_i \le \tau^1 \middle| \forall_{i\in\{1,\ldots,n\}} x_i \le \tau^1\right] \\
&\ge \Pr\left[\max_{i=1}^{n}(x_i) > \tau^1\right] + \Pr\left[\max_{i=1}^{n}(x_i) \le \tau^1\right] \times \Pr\left[\exists_{i\in\{1,\ldots,n\}}\tau^0 \le x_i \le \tau^1\right] \\
&= \lambda + (1-\lambda)\Pr\left[\exists_{i\in\{1,\ldots,n\}}\tau^0 \le x_i \le \tau^1\right] \\
&= \lambda + (1-\lambda)\Pr\left[\exists_{i\in\{1,\ldots,n\}} y_i = 1\right].
\end{aligned}$$

This implies

$$\Pr\left[\exists_{i\in\{1,\ldots,n\}} y_i = 1\right] \le \frac{\rho}{1-\lambda}.$$

□

For an arbitrary index i, the following lemma upper bounds the probability that x_i is within the range $[\tau^0, \tau^1]$. Later, we use this to show a concentration bound in Lemma 25.

Lemma 22 *Consider arbitrary numbers $\rho \in (0,1)$ and $\lambda \in [0, 1-\rho)$. Let τ^0 and τ^1 be such that $\Pr\left[\max_{i=1}^{n}(x_i) \le \tau^0\right] = 1-(\lambda+\rho)$ and $\Pr\left[\max_{i=1}^{n}(x_i) \le \tau^1\right] = 1-\lambda$. Let y_i be a random binary variable that is 1 if $\tau^0 \le x_i \le \tau^1$ and 0 otherwise. Assuming the no ε-superstars assumption we have*

$$\Pr[y_j = 1] \le \frac{\Pr[j = \arg\max_{i=1}^{n} x_i]}{1-(\lambda+\rho)} \le \frac{\varepsilon}{1-(\lambda+\rho)}.$$

Proof : For any j we have

$$\begin{aligned}
\Pr\left[j = \arg\max_{i=1}^{n} x_i\right] &\ge \Pr\left[x_j \ge \tau^0\right]\Pr\left[\arg\max_{i\in\{0,\ldots,n\}\setminus j} x_i < \tau^0\right] \\
&\ge \Pr\left[x_j \ge \tau^0\right]\Pr\left[\arg\max_{i\in\{0,\ldots,n\}} x_i < \tau^0\right] \\
&= \Pr\left[x_j \ge \tau^0\right]\left(1-(\lambda+\rho)\right) \\
&\ge \Pr\left[\tau^1 \ge x_j \ge \tau^0\right]\left(1-(\lambda+\rho)\right) \\
&= \Pr[y_j = 1]\left(1-(\lambda+\rho)\right).
\end{aligned}$$

This together with the no-superstars assumption implies that

$$\Pr[y_j = 1] \le \frac{\Pr[j = \arg\max_{i=1}^{n} x_i]}{1-(\lambda+\rho)} \le \frac{\varepsilon}{1-(\lambda+\rho)}.$$

□

The following lemma, for an arbitrary set S of indices, compares the expected number of x_i's that are in a range $[\tau^0, \tau^1]$ with the probability of observing at least one x_i in the range $[\tau^0, \tau^1]$. We later use this to exchange $\Pr\left[\exists_{i\in S} x_i \in [\tau^0, \tau^1]\right]$ and $\sum_{i\in S}\Pr\left[x_i \in [\tau^0, \tau^1]\right]$.

Lemma 23 *Consider arbitrary numbers $\rho \in (0,1)$ and $\lambda \in [0, 1-\rho)$. Let τ^0 and τ^1 be such that $\Pr\left[\max_{i=1}^{n}(x_i) \le \tau^0\right] = 1-(\lambda+\rho)$ and $\Pr\left[\max_{i=1}^{n}(x_i) \le \tau^1\right] = 1-\lambda$. Let y_i be a random binary variable that is 1 if $\tau^0 \le x_i \le \tau^1$ and 0 otherwise. Let $p'_i = \Pr[y_i = 1]$. For any set $S \subseteq \{1, \ldots, n\}$ we have*

$$\max\left(1 - \sum_{i\in S} p'_i, 1 - \frac{\rho}{1-\lambda}\right)\sum_{i\in S} p'_i \le \Pr[\exists_{i\in S} y_i = 1] \le \sum_{i\in S} p'_i.$$

Proof : We have

$$\begin{aligned}\Pr\left[\exists_{i\in S} y_i = 1\right] &= 1 - \Pr\left[\forall_{i\in S} y_i = 0\right]\\ &= 1 - \Pi_{i\in S}(1 - p'_i)\\ &\geq 1 - \exp\Big(-\sum_{i\in S} p'_i\Big).\end{aligned}$$

This implies that

$$\begin{aligned}\sum_{i\in S} p'_i &\leq \log\Big(\frac{1}{1-\Pr\left[\exists_{i\in S} y_i = 1\right]}\Big)\\ &\leq \frac{1}{1-\Pr\left[\exists_{i\in S} y_i = 1\right]} - 1 && \log(\xi) \leq \xi - 1\\ &= \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\Pr\left[\exists_{i\in S} y_i = 1\right]}\\ &\leq \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\Pr\left[\exists_{i\in\{1,\dots,n\}} y_i = 1\right]}\\ &\leq \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\frac{\rho}{1-\lambda}}. && \text{Using Lemma 21}\end{aligned}$$

This implies

$$\Big(1 - \frac{\rho}{1-\lambda}\Big)\sum_{i\in S} p'_i \leq \Pr\left[\exists_{i\in S} y_i = 1\right].$$

Similarly, we have

$$\begin{aligned}\sum_{i\in S} p'_i &\leq \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\Pr\left[\exists_{i\in S} y_i = 1\right]}\\ &\leq \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\mathrm{E}\left[\sum_{i\in S} Y_i\right]}\\ &= \frac{\Pr\left[\exists_{i\in S} y_i = 1\right]}{1-\sum_{i\in S} p'_i},\end{aligned}$$

which implies

$$\Big(1 - \sum_{i\in S} p'_i\Big)\sum_{i\in S} p'_i \leq \Pr\left[\exists_{i\in S} y_i = 1\right].$$

On the other hand we have

$$\Pr\left[\exists_{i\in S} y_i = 1\right] \leq \mathrm{E}\left[\sum_{i\in S} y_i\right] = \sum_{i\in S} p'_i.$$

□

In Lemma 25 below we show the concentration of $\sum_{i\in S}\Pr\left[x_i \in [\tau^0, \tau^1]\right]$ for a set S chosen uniformly at random without replacement. To prove Lemma 25 we use a variation of Massart's inequality for sampling without replacement Van Der Vaart and Wellner (1996). Then to apply Massart's bound to $\sum_{i\in S}\Pr\left[x_i \in [\tau^0, \tau^1]\right]$, we use Lemma 22 to upper bound $\Pr\left[x_i \in [\tau^0, \tau^1]\right]$ and use Lemma 21 to lower bound $\mathrm{E}\left[\sum_{i\in S}\Pr\left[x_i \in [\tau^0, \tau^1]\right]\right]$.

Lemma 24 (Massart's inequality) *Let $\Psi_1, \dots, \Psi_n$ be a set of n numbers and let $\psi_1, \dots, \psi_c$ be a subset of $\Psi_1, \dots, \Psi_n$ drawn uniformly at random without replacement. We have*

$$\Pr\left[\Big|\frac{1}{c}\sum_{i=1}^{c}\psi_i - \bar{\Psi}\Big| \geq \gamma\right] \leq 2\exp\Big(-\frac{c^2\gamma^2}{\sum_{i=1}^{n}(\Psi_i - \bar{\Psi})^2}\Big),$$

where $\bar{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\Psi_i$, and n is assumed to be divisible by c.

Now we are ready to prove Lemma 25.

Lemma 25 *Consider arbitrary numbers $\rho, \gamma, \varepsilon, q \in (0,1)$, $\lambda \in [0, 1-\rho)$. Let S be a set of size qn, chosen uniformly at random without replacement from $D_1, \ldots, D_n$. Let τ^0 and τ^1 be such that $\Pr\left[\max_{i=1}^n(x_i) \leq \tau^0\right] = 1-(\lambda+\rho)$ and $\Pr\left[\max_{i=1}^n(x_i) \leq \tau^1\right] = 1-\lambda$. Let y_i be a random binary variable that is 1 if $\tau^0 \leq x_i \leq \tau^1$ and 0 otherwise. Let $p'_i = \Pr[y_i = 1]$. Assuming the no ε-superstars assumption, with probability $1 - 2\exp\Big(-\frac{\gamma^2 q^2 \rho(1-(\lambda+\rho))}{2\varepsilon}\Big)$ we have*

$$(1-\gamma)q\sum_{i=1}^n p'_i \leq \sum_{i\in S} p'_i \leq (1+\gamma)q\sum_{i=1}^n p'_i$$

Proof : Let z_i be a random variable that is 1 when $i \in S$ and 0 otherwise. We have

$$\begin{aligned}
\sum_{i=0}^n p'_i &\geq \Pr\left[\exists_{i\in S} y_i = 1\right] && \text{By Lemma 23}\\
&\geq \rho. && \text{By Lemma 21}
\end{aligned}$$

Moreover, by Lemma 22 we have $0 \leq p'_i \leq \frac{\varepsilon}{1-(\lambda+\rho)}$. Thus,

$$\begin{aligned}
&\Pr\left[\Big|\sum_{i\in S} p'_i - q\sum_{i=1}^n p'_i\Big| \geq \gamma q \sum_{i=1}^n p'_i\right] =\\
&\Pr\left[\Big|\frac{1}{qn}\sum_{i\in S} p'_i - \frac{1}{n}\sum_{i=1}^n p'_i\Big| \geq \gamma \frac{1}{n}\sum_{i=1}^n p'_i\right] = && \text{Multiply both sides by } \frac{1}{qn}\\
&2\exp\Big(-\frac{(qn)^2\big(\gamma\frac{1}{n}\sum_{i=1}^n p'_i\big)^2}{\sum_{i=1}^n\big(p'_i - \frac{1}{n}\sum_{i=1}^n p'_i\big)^2}\Big) = && \text{Massart bound}\\
&2\exp\Big(-q^2\gamma^2\frac{\big(\sum_{i=1}^n p'_i\big)^2}{\sum_{i=1}^n\big(p'_i - \frac{1}{n}\sum_{i=1}^n p'_i\big)^2}\Big) \leq\\
&2\exp\Big(-q^2\gamma^2\frac{\big(\sum_{i=1}^n p'_i\big)^2}{\sum_{i=1}^n {p'_i}^2 + \sum_{i=1}^n\big(\frac{1}{n}\sum_{i=1}^n p'_i\big)^2}\Big) \leq\\
&2\exp\Big(-q^2\gamma^2\frac{\big(\sum_{i=1}^n p'_i\big)^2}{2\sum_{i=1}^n {p'_i}^2}\Big) \leq\\
&2\exp\Big(-q^2\gamma^2\frac{\big(\sum_{i=1}^n p'_i\big)^2}{2\frac{\varepsilon}{1-(\lambda+\rho)}\sum_{i=1}^n p'_i}\Big) = && p'_i \leq \frac{\varepsilon}{1-(\lambda+\rho)}\\
&2\exp\Big(-\frac{q^2\gamma^2(1-(\lambda+\rho))}{2\varepsilon}\sum_{i=1}^n p'_i\Big) \leq\\
&2\exp\Big(-\frac{\gamma^2 q^2\rho(1-(\lambda+\rho))}{2\varepsilon}\Big) && \sum_{i=1}^n p'_i \geq \rho
\end{aligned}$$

□

Next, we use Lemma 25 together with Lemma 23 to show the concentration of $\Pr\left[\exists_{i\in S} x_i \in [\tau^0, \tau^1]\right]$ for a set S chosen uniformly at random without replacement.

Lemma 26 *Consider arbitrary numbers $\rho, \gamma, \varepsilon, q \in (0,1)$, $\lambda \in [0, 1-\rho)$. Let S be a set of size qn, chosen uniformly at random without replacement from $D_1, \ldots, D_n$. Let τ^0 and τ^1 be such that $\Pr\left[\max_{i=1}^n(x_i) \leq \tau^0\right] =$*

$1-(\lambda+\rho)$ *and* $\Pr\left[\max_{i=1}^{n}(x_i) \leq \tau^1\right] = 1-\lambda$. *Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \leq x_i \leq \tau^1$ *and* 0 *otherwise. Let* $p'_i = \Pr[y_i = 1]$. *Assuming the no* ε*-superstars assumption, with probability* $1-2\exp\Big(-\frac{\gamma^2 q^2 \rho(1-(\lambda+\rho))}{2\varepsilon}\Big)$ *we have*

$$\big(1-\gamma-\frac{2q\rho}{1-(\lambda+\rho)}\big)q\sum_{i=1}^{n} p'_i \leq \Pr[\exists_{i\in S} y_i = 1] \leq (1+\gamma)q\sum_{i=1}^{n} p'_i$$

Proof : With probability $1-2\exp\Big(-\frac{\gamma^2 q^2 \rho(1-(\lambda+\rho))}{2\varepsilon}\Big)$ Lemma 25 holds and we have

$$\begin{aligned}
\Pr[\exists_{i\in S} y_i = 1] &\leq \sum_{i\in S} p'_i && \text{By Lemma 23}\\
&\leq (1+\gamma)q\sum_{i=1}^{n} p'_i. && \text{By Lemma 25}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\Pr[\exists_{i\in S} y_i = 1] &\geq \big(1-\sum_{i\in S} p'_i\big)\sum_{i\in S} p'_i && \text{By Lemma 23}\\
&\geq \big(1-\sum_{i\in S} p'_i\big)(1-\gamma)q\sum_{i=1}^{n} p'_i && \text{By Lemma 25}\\
&\geq \big(1-(1+\gamma)q\sum_{i=1}^{n} p'_i\big)(1-\gamma)q\sum_{i=1}^{n} p'_i && \text{By Lemma 25}\\
&\geq \big(1-(1+\gamma)q\frac{1-\lambda}{1-(\lambda+\rho)}\Pr[\exists_{i\in S} y_i = 1]\big)(1-\gamma)q\sum_{i=1}^{n} p'_i && \text{By Lemma 23}\\
&\geq \big(1-2q\frac{1-\lambda}{1-(\lambda+\rho)}\Pr[\exists_{i\in S} y_i = 1]\big)(1-\gamma)q\sum_{i=1}^{n} p'_i\\
&\geq \big(1-2q\frac{1-\lambda}{1-(\lambda+\rho)}\frac{\rho}{1-\lambda}\big)(1-\gamma)q\sum_{i=1}^{n} p'_i && \text{By Lemma 21}\\
&\geq \big(1-\frac{2q\rho}{1-(\lambda+\rho)}\big)(1-\gamma)q\sum_{i=1}^{n} p'_i\\
&\geq \big(1-\gamma-\frac{2q\rho}{1-(\lambda+\rho)}\big)q\sum_{i=1}^{n} p'_i.
\end{aligned}$$

□

The following corollary is a simplified (and restricted) variation of Lemma 26.

Corollary 27 *Consider arbitrary numbers* $\rho, \lambda_0, \gamma, \delta \in (0,1)$, $\lambda \in [0, 1-(\lambda_0+\rho)]$ *and* $q \in \big(0, \min\big(\frac{\gamma\lambda_0}{2\rho}, 1\big)\big)$. *Set* $\varepsilon = \frac{\gamma^2 q^2 \rho\lambda_0}{2\log\frac{2}{\delta}}$. *Let* S *be a set of size* qn, *chosen uniformly at random without replacement from* $D_1, \ldots, D_n$. *Let* τ^0 *and* τ^1 *be such that* $\Pr\left[\max_{i=1}^{n}(x_i) \leq \tau^0\right] = 1-(\lambda+\rho)$ *and* $\Pr\left[\max_{i=1}^{n}(x_i) \leq \tau^1\right] = 1-\lambda$. *Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \leq x_i \leq \tau^1$ *and* 0 *otherwise. Let* $p'_i = \Pr[y_i = 1]$. *Assuming the no* ε*-superstars assumption, with probability* $1-\delta$ *we have*

$$(1-2\gamma)q\sum_{i=1}^{n} p'_i \leq \Pr[\exists_{i\in S} y_i = 1] \leq (1+\gamma)q\sum_{i=1}^{n} p'_i$$

Proof : Note that Lemma 26 holds with probability

$$\begin{aligned}
1-2\exp\Big(-\frac{\gamma^2 q^2\rho(1-(\lambda+\rho))}{2\varepsilon}\Big) &\geq 1-2\exp\Big(-\frac{\gamma^2 q^2\rho\lambda_0}{2\varepsilon}\Big)\\
&= 1-2\exp\left(-\frac{\gamma^2 q^2\rho\lambda_0}{2\big(\frac{\gamma^2 q^2\rho\lambda_0}{2\log\frac{2}{\delta}}\big)}\right)\\
&= 1-2\exp\big(-\log\frac{2}{\delta}\big)\\
&= 1-\delta. \qquad (4)
\end{aligned}$$

Note that Lemma 26 directly gives us $\Pr[\exists_{i\in S} y_i = 1] \leq (1+\gamma)q\sum_{i=1}^{n} p'_i$. Moreover, we have

$$\begin{aligned}
\Pr[\exists_{i\in S} y_i = 1] &\geq \big(1-\gamma-\frac{2q\rho}{1-(\lambda+\rho)}\big)q\sum_{i=1}^{n} p'_i \qquad \text{Lemma 26}\\
&\geq \big(1-\gamma-\frac{2q\rho}{\lambda_0}\big)q\sum_{i=1}^{n} p'_i\\
&\geq \big(1-\gamma-\frac{2\frac{\gamma\lambda_0}{2\rho}\rho}{\lambda_0}\big)q\sum_{i=1}^{n} p'_i\\
&= (1-2\gamma)q\sum_{i=1}^{n} p'_i.
\end{aligned}$$

□

We can now prove Lemma 11. We will restate it as Lemma 28 below for convenience. Recall that for the purpose of this lemma for some arbitrary $\gamma \in (0,1)$ we set $\lambda_0 = \gamma$, $\rho = \gamma^3$, $q = \frac{\gamma^2}{2}$, and $\delta = \frac{\gamma^6}{4}$.

Lemma 28 *Let S be a set of size qn, chosen uniformly at random without replacement from $D_1, \ldots, D_n$. With probability $1-\frac{\gamma^3}{2}$ for all $\zeta \in \{0,\ldots,c-1\}$ we have*

$$(1-3\gamma)q\sum_{i=1}^{n} p_i^{\zeta} \leq \Pr[t_\zeta \leq x_S < t_{\zeta+1}] \leq (1+\gamma)q\sum_{i=1}^{n} p_i^{\zeta},$$

where $p_i^{\zeta} = \Pr[t_\zeta \leq x_i < t_{\zeta+1}]$, assuming the no ε-superstars assumption with $\varepsilon = \frac{\gamma^{10}}{24\log(\frac{2}{\gamma^2})}$.

Proof : Note that by Corollary 27, for a fixed $\zeta \in \{0,\ldots,c-1\}$ with probability $1-\delta = 1-\frac{\gamma^6}{4}$ we have

$$(1-2\gamma)q\sum_{i=1}^{n} p_i^{\zeta} \leq \Pr[\exists_{i\in S} t_\zeta \leq x_i < t_{\zeta+1}] \leq (1+\gamma)q\sum_{i=1}^{n} p_i^{\zeta}. \qquad (5)$$

By the union bound, this holds for all $\zeta \in \{0,\ldots,c-1\}$ with probability at least

$$1-c\frac{\gamma^6}{4} = 1-\frac{1-\lambda_0}{\rho}\frac{\gamma^6}{4} \geq 1-\frac{\gamma^6}{4\rho} = 1-\frac{\gamma^3}{4}.$$

Similarly, using Lemma 16 with probability at least $1-\frac{\gamma^3}{4}$ for all $\zeta \in \{0,\ldots,c-1\}$ we have

$$\Pr[\exists_{i\in S} t_{\zeta+1} \leq x_i] \leq \frac{2q}{\lambda_0} = \gamma. \qquad (6)$$

Next, we prove the statement of the lemma assuming that for all $\zeta \in \{0,\ldots,c-1\}$ Inequalities 5 and 6 hold. First note that we have

$$\begin{aligned}
\Pr[t_\zeta \leq x_S < t_{\zeta+1}] &\leq \Pr[\exists_{i\in S} t_\zeta \leq x_i < t_{\zeta+1}]\\
&\leq (1+\gamma)q\sum_{i=1}^{n} p_i^{\zeta}. \qquad \text{By Inequality 5}
\end{aligned}$$

This proves the upper bound. On the other hand we have

$$\begin{aligned}
\Pr[t_\zeta \le x_S < t_{\zeta+1}] &\ge \Pr[\nexists_{i\in S} x_i \ge t_{\zeta+1}] \times \Pr[\exists_{i\in S} t_\zeta \le x_i < t_{\zeta+1}] \\
&= \Big(1 - \Pr[\exists_{i\in S} x_i \ge t_{\zeta+1}]\Big) \times \Pr[\exists_{i\in S} t_\zeta \le x_i < t_{\zeta+1}] \\
&\ge (1-\gamma)\Pr[\exists_{i\in S} t_\zeta \le x_i < t_{\zeta+1}] && \text{By Inequality 6} \\
&\ge (1-\gamma)(1-2\gamma) q \sum_{i=1}^{n} p_i^\zeta && \text{By Inequality 5} \\
&\ge (1-3\gamma) q \sum_{i=1}^{n} p_i^\zeta .
\end{aligned}$$

□

The following technical lemma will be useful for proving Lemma 16 and Lemma 17.

Lemma 29 *Let $\chi_1, \dots, \chi_m$ be a sequence of independent binary random variables. We have*

$$\Pr\left[\sum_{i=1}^{m} \chi_i \ge 2\right] \le \Pr[\exists_i \chi_i = 1]^2 .$$

Proof : We have

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{m} \chi_i \ge 2\right] &= \sum_{j=1}^{m} \left(\Pr[\forall_{i<j} \chi_i = 0] \Pr[\chi_j = 1] \Pr\left[\sum_{i=j+1}^{m} \chi_i \ge 1\right] \right) \\
&\le \sum_{j=1}^{m} \left(\Pr[\forall_{i<j} \chi_i = 0] \Pr[\chi_j = 1] \Pr\left[\sum_{i=0}^{m} \chi_i \ge 1\right] \right) \\
&= \Pr\left[\sum_{i=0}^{m} \chi_i \ge 1\right] \sum_{j=1}^{m} \Big(\Pr[\forall_{i<j} \chi_i = 0] \Pr[\chi_j = 1] \Big) \\
&= \Pr\left[\sum_{i=0}^{m} \chi_i \ge 1\right]^2 \\
&= \Pr[\exists_i \chi_i = 1]^2 .
\end{aligned}$$

□

We can now prove Lemma 16. For a small set of indices S chosen uniformly at random, we wish to upper bound the probability of observing at least two x_i's with $i \in S$ above a threshold τ^0. We declare this as a failure case in our algorithm in subsection C. For convenience we restate as Lemma 30 below.

Lemma 30 *Consider arbitrary numbers $\lambda_0, \gamma, \delta, q \in (0,1)$, $\rho \in (0, 1-\lambda_0)$. Set $\varepsilon = \frac{\gamma^2 q^2 \rho \lambda_0}{2 \log \frac{2}{\delta}}$. Let S be a set of size qn, chosen uniformly at random without replacement from $D_1, \dots, D_n$. Let τ^0 be such that $\Pr\left[\max_{i=1}^{n}(x_i) \le \tau^0\right] = 1 - \rho$. Let y_i be a random binary variable that is 1 if $\tau^0 \le x_i$ and 0 otherwise. Let $p'_i = \Pr[y_i = 1]$. Assuming the no ε-superstars assumption, with probability $1-\delta$ we have*

$$\Pr[\exists_{i\in S} y_i = 1] \le \frac{2q}{\lambda_0}$$

and

$$\Pr\left[\sum_{i\in S} y_i \ge 2\right] \le \frac{4q^2}{\lambda_0^2} .$$

Proof : With probability $1-2\exp\left(-\frac{\gamma^2 q^2 \rho(1-(\lambda+\rho))}{2\varepsilon}\right) \geq 1-\delta$ (see Inequality 4), Lemma 25 holds and hence we have

$$\begin{aligned}
\Pr\left[\exists_{i\in S} y_i = 1\right] &\leq \sum_{i\in S} p'_i && \text{By Lemma 23} \\
&\leq 2q \sum_{i=1}^{n} p'_i && \text{By Lemma 25 with } \gamma < 1 \\
&\leq 2q \sum_{i=1}^{n} \frac{\Pr\left[i = \arg\max_{j=1}^{n} x_j\right]}{\lambda_0} && \text{By Lemma 22 with } \lambda = 0, \rho = 1-\lambda_0 \\
&\leq \frac{2q}{\lambda_0}. && \sum_{i=1}^{n} \Pr\left[i = \arg\max_{j=1}^{n} x_j\right] = 1
\end{aligned}$$

and hence, we have

$$\begin{aligned}
\Pr\left[\sum_{i\in S} y_i \geq 2\right] &\leq \Pr\left[\exists_{i\in S} y_i = 1\right]^2 && \text{By Lemma 29} \\
&\leq \frac{4q^2}{\lambda_0^2}.
\end{aligned}$$

□

Finally we will prove Lemma 17. We wish to upper bound the probability of observing at least two x_i's within a narrow range $[\tau^0, \tau^1]$. We declare this as a failure case in our algorithm in subsection C. For convenience we restate as Lemma 31 below.

Lemma 31 *Consider arbitrary numbers* $\rho, \lambda_0 \in (0,1)$ *and* $\lambda \in [0, 1-(\lambda_0+\rho)]$. *Let* τ^0 *and* τ^1 *be such that* $\Pr\left[\max_{i=1}^{n}(x_i) \leq \tau^0\right] = 1-(\lambda+\rho)$ *and* $\Pr\left[\max_{i=1}^{n}(x_i) \leq \tau^1\right] = 1-\lambda$. *Let* y_i *be a random binary variable that is* 1 *if* $\tau^0 \leq x_i \leq \tau^1$ *and* 0 *otherwise. We have*

$$\Pr\left[\sum_{i=1}^{n} y_i \geq 2\right] \leq \frac{\rho^2}{\lambda_0^2}.$$

Proof : We have

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} y_i \geq 2\right] &\leq \Pr\left[\exists_{i=1}^{n} y_i = 1\right]^2 && \text{By Lemma 29} \\
&\leq \left(\frac{\rho}{1-\lambda}\right) && \text{By Lemma 21} \\
&\leq \frac{\rho^2}{\lambda_0^2}.
\end{aligned}$$

□