

# Online Resource Allocation with Samples

Negin Golrezaei

Sloan School of Management, Massachusetts Institute of Technology, golrezaei@mit.edu

Patrick Jaillet

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, jaillet@mit.edu

Zijie Zhou

Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, 02139, zhou98@mit.edu

We study an online resource allocation problem under uncertainty about demand and about the reward of each type of demand (agents) for the resource. Even though dealing with demand uncertainty in resource allocation problems has been the topic of many papers in the literature, the challenge of not knowing rewards has been barely explored. The lack of knowledge about agents' rewards is inspired by the problem of allocating units of a new resource (e.g., newly developed vaccines or drugs) with unknown effectiveness/value. For such settings, we assume that we can *test* the market before the allocation period starts. During the test period, we sample each agent in the market with probability  $p$ . We study how to optimally exploit the *sample information* in our online resource allocation problem under adversarial arrival processes. We present an asymptotically optimal algorithm that achieves  $1 - \Theta(1/(p\sqrt{m}))$  competitive ratio, where  $m$  is the number of available units of the resource. By characterizing an upper bound on the competitive ratio of any randomized and deterministic algorithm, we show that our competitive ratio of  $1 - \Theta(1/(p\sqrt{m}))$  is tight for any  $p = \omega(1/\sqrt{m})$ . That asymptotic optimality is possible with sample information highlights the significant advantage of running a test period for new resources. We demonstrate the efficacy of our proposed algorithm using a dataset that contains the number of COVID-19 related hospitalized patients across different age groups.

*Key words:* online resource allocation, sample information, new resources, competitive ratio, COVID-19

---

## 1. Introduction

In online resource allocation problems, the goal is to allocate a limited number of a given resource to heterogeneous demand/agents that arrive over time. These problems are notoriously challenging mainly because of the demand uncertainty and scarcity of the resource. Such problems get even more challenging for newly developed resources (e.g., new drugs, products, and services). For such a resource, the effectiveness/value of the resource for different types of agents may not be fully known. To overcome this additional challenge, businesses, for example, offer free products in an exchange for honest feedback (productreviewmom.com 2022), and pharmaceutical companies test potential treatments/drugs in

human volunteers (Pfizer 2022). These practices raise the following key question: *can and to what extent such feedback improve the efficiency of online resource allocation?*

To answer this question, we consider a decision-maker who aims to allocate her available resource to two types of unit-demand agents with unknown (expected) rewards, where type 1 has a higher expected reward than type 2.<sup>1</sup> The total number of agents (i.e., the market size), as well as, the number of agents of type 1 and 2, denoted by  $h$  and  $\ell$  respectively, are chosen adversarially, and hence are unknown to the decision-maker. Before the allocation period starts, the decision-maker tests the market by, for example, making a public announcement and offering resources for free. We assume that with probability  $p \in (0, 1)$ , each of the  $h + \ell$  agents sees and reacts to the announcement,<sup>2</sup> and gets one unit of the resource, where we assume that  $p$  is known to the decision-maker. (We will discuss this assumption later in this section.) These agents then provide feedback on their realized reward for the resource in return. That is, we assume that all the agents in the test period will get a resource. As we will show in Section 7.1, this assumption can be relaxed by limiting the number of available resources during the test period. The test procedure supplies the decision-maker with some information about agents' expected rewards, as well as, the size of the market for each type of agent. We refer to this information as *sample information*.

After the test period ends, the remaining agents arrive over time according to an adversarially-chosen order. For each arriving agent, the decision-maker has to make an irrevocable decision about accepting him and allocating him one unit of the resource or rejecting him. The decision-maker who has  $m$  units of the resource when the allocation period starts makes acceptance/rejection decisions while being uncertain about the number, type of future agents, and their expected rewards. The decision-maker is also uncertain about which types of agents earn higher rewards upon receiving the resource. For such a demanding setting, our goal is to design efficient resource allocation algorithms that can optimally utilize the sample information under any possible arrival sequence. In other words, we measure the performance of the algorithm in terms of its competitive ratio, which is the expected ratio of the reward of the algorithm to the reward of the optimal

<sup>1</sup> In Section 7.4, we discuss the case of having more than two types of agents.

<sup>2</sup> In Section 7.3, we allow the probability  $p$  to be different for type 1 and 2 agents.

clairvoyant solution that knows the arrival sequence and the expected reward of agents in advance; see Section 2 for the model and the formal definition of the competitive ratio.

Before presenting our contributions, we make two remarks about our model. First, our model bears resemblance to the proposed models in Correa et al. (2021), Kaplan et al. (2022) for secretary and online bipartite matching problems, respectively. In Correa et al. (2021), each of the secretaries is placed in a sample set with probability  $p$ , where the value of the sampled secretaries will be disclosed to the decision-maker before the decision period starts. In Kaplan et al. (2022) that studies an online bipartite matching problem, each agent independently will be placed in a sample set with probability  $p$ . While at a high level, these works seek to design algorithms that can take advantage of samples, the nature of their considered problems is different from ours, preventing us to use their designed algorithms for our setting. We discuss the details in Section 1.2.

Second, our model is a special case of the single-leg revenue management problem, which has been widely studied in the literature; see, e.g., Littlewood (1972), Amaruchkul et al. (2007), Ball and Queyranne (2009), Gallego et al. (2009), Ferreira et al. (2018), Jasin (2015), Hwang et al. (2021), Golrezaei and Yao (2021). In all of these aforementioned works, while the decision-maker may be uncertain about the demand (i.e., the number and the order of the arrivals), the obtained reward of different types of demand upon receiving the resource is fully known to the decision-maker. This is in sharp contrast with our model in which these rewards are not known to the decision-maker, adding extra complexity to our problem. (See also Section 1.2 for a discussion about previous works on revenue management problems with limited demand information, but full knowledge of rewards.)

### 1.1. Our Contributions

In addition to our modeling contribution, our work makes the following contributions.

**Impossibility results.** To shed light on the value of sample information, in Section 3, we consider alternative scenarios under which either the sample information is not available or the sample information is not very informative due to the lack of some additional knowledge (e.g., the sampling probability). In all of the considered scenarios, we show that it is not possible to design asymptotically optimal algorithms whose competitive ratio goes to one as the number of resources  $m$  increases. While in the first scenario, no sample information is not available (i.e., the *sampling probability*  $p = 0$ ), in the second scenario, the sample information is available but the sampling probability  $p$  is not known to the

decision-maker. The result for the second scenario justifies our assumption about knowing the sampling probability; see (Correa et al. 2021) for a similar assumption. Nonetheless, this assumption can be also justified by the fact that the outreach program is designed by the decision-maker herself, and hence  $p$  can be well estimated using historical data for similar outreach programs. See also our case study in Section 6 where we examine the robustness of our proposed algorithm (which we will present shortly) to the lack of knowledge of the sampling probability.<sup>3</sup>

**Asymptotically optimal protection level algorithm.** In light of our impossibility results in Section 3, we design a simple, yet effective protection level algorithm that optimally utilizes the sample information; see Algorithm 1 in Section 4. The algorithm uses the sample information to obtain an estimate of the expected reward of each type of agents. If the estimated reward of type 1 is greater than that of type 2, the algorithm *protects* type 1 agents, otherwise type 2 agents will be protected. In each of these cases, the algorithm uses the sample information to estimate the protection level for the type that has a higher estimated average reward.

In Theorem 1, we present the competitive ratio of our proposed algorithm for any finite value of  $m$ . In addition, in Proposition 2, we show that our algorithm is asymptotically optimal as  $m$  goes to infinity. More precisely, we show that for any  $p = \omega(1/\sqrt{m})$ <sup>4</sup> (which includes constant values for  $p$  that does not scale with  $m$ ), the asymptotic competitive ratio of our algorithm is on the order of  $1 - \Theta(1/(p\sqrt{m}))$ .<sup>5</sup> As we show in Section 7.1, the same asymptotic competitive ratio continues to hold even when there is a capacity constraint during the test period. (There we show that when the number of resources during the test period is in the order of  $\omega(\sqrt{m})$ , the asymptotic competitive ratio of (a slightly modified version of) Algorithm 1 is in the order of  $1 - \Theta(1/(p\sqrt{m}))$ .)

<sup>3</sup> In Appendix K, using numerical studies and under adversarial arrival processes, we further test the robustness of our proposed algorithm to the lack of knowledge of  $p$ . There we show that when the estimation error in  $p$  is 10%, the competitive ratio decreases by at most 6%. Similar results are obtained when the estimation error in  $p$  is 20%.

<sup>4</sup> We scale  $p$  with  $m$  because of the cost of testing the market. With a large value of  $m$ , if the market size turns out to be large too (e.g., proportional to  $m$ ), the decision-maker will incur a large cost during the test period, justifying our choice of  $p = \omega(1/\sqrt{m})$ . Nonetheless, our results hold for constant  $p$ 's, and theoretically speaking, handling the case of  $p = \omega(1/\sqrt{m})$  is more challenging than that of constant  $p$ 's. This is because, under constant  $p$ 's, the sample information is more informative, easing the analysis of the algorithm.

<sup>5</sup> For the case of  $p = O(\frac{1}{\sqrt{m}})$  where  $p$  goes to zero very fast as  $m$  grows, in Section J, we design another algorithm whose competitive ratio is  $1/2$ . We note that our upper bound in Section 5.2 confirms the challenges of such an extreme case for the sampling probability. As we show there, when  $p$  goes to zero very fast (i.e.,  $p = O(1/m)$ ), obtaining asymptotic optimality is impossible.

This result shows that the sample information can be extremely useful to improve the performance of resource allocation algorithms. This is because in the absence of the sample information, and even when the expected rewards of agents are fully known, as shown in Ball and Queyranne (2009), the competitive ratio of any algorithm cannot exceed  $1/(2 - r_2/r_1)$  no matter how large  $m$  is, where  $r_1 > r_2$  are the expected rewards for the two types of agents. Here, we show that we can break the barrier of  $1/(2 - r_2/r_1)$  by taking advantage of the sample information in a very challenging setting, where the expected reward of agents is not known to the decision-maker. This is mainly because the sample information can be used to infer some knowledge about the number of agents of different types in the online arrival sequence, gaining some knowledge about the adversarially-chosen arrival sequence.

That being said, due to the adversarial nature of the demand process, it is not obvious if the sample information can lead to asymptotic optimality when  $p = \omega(1/\sqrt{m})$ . Even when the sampling probability  $p$  is constant, with a large number of resources, the sample information may not be very illuminating. Consider the case where either the market size  $h$  or  $\ell$  is small. In such cases, the number of agents in the sample set for at least one of the types will be small and hence the decision-maker cannot correctly estimate the expected rewards of the agents. This, in turn, can lead to the decision-maker protecting the wrong (low-reward) type. Considering this, it is quite remarkable that our algorithm obtains asymptotic optimality even when  $p$  shrinks as a function of  $m$ . As we will discuss shortly, the competitive ratio of our algorithm is tight concerning both  $m$  and  $p$ .

We now comment on our technical contributions in characterizing the competitive ratio of our algorithm. The proof of Theorem 1 is quite involved and is divided into three main cases, where each case bounds the competitive ratio of the algorithm when the total number of agents of type 1 and 2 falls into a certain region. The most challenging region is the one in which the total number of type 1 agents (i.e.,  $h$ ) is large. In this case, the algorithm may lose reward in three aspects: not protecting type 1 agents, over-protecting type 1 agents, and under-protecting type 1 agents. Recall that in our setting, the decision-maker does not even know which type has the higher reward, and hence our algorithm may wrongfully protect type 2 agents. In addition, for large values of  $h$ , even when the right type is protected, by over- and under-protecting the protected type, our algorithm can lose some rewards. At a high level, we overcome these challenges, by showing that either (i)

the algorithm protects the right (high-reward) type with high probability, or (ii) thanks to setting protection levels using the sample information, the loss of the algorithm is small if it protects the wrong (low-reward) type.

We note that while characterizing the bound for the region with large  $h$  is the most challenging one, the proof shows that the competitive ratio is smallest when  $h$  is small. This is because, with small  $h$ , the sample information may not even reveal which type has the highest reward, leading the algorithm to protect the wrong type.

**Upper bound on the competitive ratio of any deterministic and randomized algorithms.** Our proposed algorithm obtains the asymptotic competitive ratio of  $1 - \Theta(1/(p\sqrt{m}))$  when  $p = \omega(1/\sqrt{m})$ . Such a superb performance makes us wonder if we can design an algorithm with even a better asymptotic competitive ratio. In Section 5, we present Theorems 2 and 3, which show that with  $p = \omega(1/\sqrt{m})$ , even when the decision-maker is fully aware of the expected reward of agents, no deterministic and randomized algorithms can obtain an asymptotic competitive ratio better than  $1 - \Theta(1/(p\sqrt{m}))$ . To show the upper bound of  $1 - \Theta(1/(p\sqrt{m}))$  on the competitive ratio for any deterministic algorithm, in the proof of Theorem 2, we consider a family of arrivals, wherein this family, a large number of type 2 agents (i.e., the type with a lower reward) arrive first, followed by some number of type 1 agents. Under this family of arrivals, any deterministic algorithm has to decide how many type 2 agents to accept based on the number of type 1 agents in the sample. We show that due to the lack of precise knowledge about the number of type 1 agents, no deterministic algorithm can do better than  $1 - \Theta(1/(p\sqrt{m}))$  on the constructed family of arrivals. To show the same result for any randomized algorithms, we first derive a variation of Yao’s Lemma that could be of independent interest; see Lemma 15. We then devise a distribution over the family of the arrival sequence that we considered in Theorem 2 and show the upper bound of  $1 - \Theta(1/(p\sqrt{m}))$  using Lemma 15.

For the case of  $p = O(1/\sqrt{m})$ , we present another upper bound in Section 5.2. This non-asymptotic upper bound, which is valid for any value of  $m$ , shows that when  $p = O(1/m)$ , it is not possible to exceed the bound of  $1/(2 - r_2/r_1)$ , where  $1/(2 - r_2/r_1)$  is the upper bound in Ball and Queyranne (2009) for a setting with no sample information, but known rewards. This shows that when  $p$  is very small, the sample information is not sufficient to overcome the challenge of not knowing the rewards.

**Case study.** We perform a case study in Section 6 using the “Laboratory-Confirmed COVID-19-Associated Hospitalizations” dataset, which contains the number of bi-weekly cases of COVID-19-associated hospitalizations in the US from March 7<sup>th</sup>, 2020 to February 5<sup>th</sup>, 2022. The dataset is obtained from the following website: [gis.cdc.gov/grasp/covidnet/covid19\\_5.html](https://gis.cdc.gov/grasp/covidnet/covid19_5.html). We study how to use our algorithm to allocate limited hospital resources (e.g., a certain medicine) to different types of COVID-19 patients while having access to some sample information. We show that the average competitive ratio of our algorithm in various realistic settings exceed 0.88 and our algorithm substantially outperforms a naive algorithm that does not use the sample information. Further, we observe that our algorithm maintains its performance when it only has access to an estimate of the sampling probability  $p$ .

## 1.2. Other Related Works

Our work is related to various streams of works in the literature.

**Online decision-making with samples.** As stated earlier, our model is related to some of the recent works on online decision-making with samples. Correa et al. (2021) study the secretary problem under both adversarial and random order arrival models with the independent sampling process, where in the secretary problem, the goal is to maximize the probability of selecting the best applicant. For the adversarial model, with the knowledge of the true sample information, they design a simple threshold-based algorithm (with the threshold of  $k = 1/(1 - p)$ ) that achieves a competitive ratio of  $kp^k(1 - p)$ . They show that the bound of  $kp^k(1 - p)$  is tight. (Similar results are obtained for the random order arrival models.) Kaplan et al. (2022) study an online weighted bipartite matching problem under an adversarial arrival with a similar independent sampling process. They analyze a simple greedy algorithm which does not depend on  $p$  and show that it achieves a competitive ratio of  $p(1 - p)$  for  $p \leq 1/2$  and  $1/4$  for  $p > 1/2$ . We note that without samples, the best competitive ratio for the adversarial secretary problem and the online matching problem in Kaplan et al. (2022) is zero. Thus, similar to our work, the aforementioned works show that the sample information can be very helpful in enabling more efficient online decision-making.

**Revenue management with limited demand information.** In practice, it is hard to fully predict demand, and hence revenue management with limited demand information becomes an essential problem. Lan et al. (2008) study a single-leg revenue management

problem with predicted lower and upper bounds of the demand for each type of agent and present an optimal algorithm. Perakis and Roels (2010) study a network revenue management problem with lower and upper bounds on demand of each type of agent and design an approximation algorithm.<sup>6</sup> Besides the quantity-based revenue management problem discussed above, some works study the price-based revenue management problem with limited demand information. For example, Besbes and Zeevi (2012) design an algorithm that dynamically adjusts prices to maximize rewards under the model where the demand at each point in time is determined by the price. Araman and Caldentey (2011) study a price-based revenue management problem with a parametric demand model (e.g., linear demand model with unknown coefficients). See also Bu et al. (2020) for a work that considers the problem of learning optimal price while having access to some offline demand data under certain prices. Our work contributes to this literature by considering a novel and practical adversarial model with samples, where the samples provide limited demand information.

**Online algorithm design with machine learned advice.** The sample information in our setting provides some information to the decision-maker regarding the adversarially-chosen arrival sequence, allowing us to significantly improve the worst-case guarantee of our algorithm. Improving the worst-case guarantee of online algorithms with the help of extra information has been the topic of recent literature on algorithm design with machine-learned advice. See, for example, Antoniadis et al. (2020) for using advice on the maximum value of secretaries in the online secretary problems, Lattanzi et al. (2020) for using advice on the weights of jobs in online scheduling problems, Lykouris and Vassilvtiskii (2018) for using the advice in the online caching problem, Balseiro et al. (2022) for using advice in single leg revenue management problems,<sup>7</sup> and Jin and Ma (2022) for using advice in online bipartite matching problems. Our work contributes to this line of work by presenting the first algorithm that optimally exploits rather unstructured advice obtained through the sample information in an online resource allocation problem.

**Multi-armed bandits.** Our setting is also related to the vast literature on multi-armed bandits; see, for example, Thompson (1933), Auer et al. (2002a,b), Balseiro et al. (2019),

<sup>6</sup> See also Hwang et al. (2021), Esfandiari et al. (2015) for works that consider a hybrid arrival model with both adversarial and stochastic components. For such models, the stochastic component can reduce the uncertainty in the demand process.

<sup>7</sup> In Balseiro et al. (2022), the advice is a predicted demand vector.



Van Parys and Golrezaei (2020), Niazadeh et al. (2021), Chen and Yang (2022), Chen et al. (2021b). In this literature, it is assumed that there are some arms/options with unknown expected rewards, and the goal is to identify the best arm by suffering from minimal regret, where the regret is computed against a clairvoyant policy that knows the expected reward of arms in advance. In our setting, similar to the bandit settings, the rewards of agents are unknown. However, unlike the bandit setting, the algorithm aims to partially learn these rewards via the sample information, rather than the feedback it receives throughout the allocation period. Our upper bound results, which are obtained under the setting when the rewards are fully known, show that using the feedback during the allocation period does not improve the asymptotic competitive ratio of our algorithm. This is because, in our setting, the order of arrivals and the number of agents of different types are chosen adversarially. Hence, in the worst case, the feedback throughout the allocation periods does not add any value, as the decision-maker is not able to acquire the right feedback at the right time. What feedback the decision-maker can receive is mainly determined by the adversary.

**Online resource allocation.** Our work is also related to the rich literature on online resource allocation. Devanur and Hayes (2009), Feldman et al. (2010), Agrawal et al. (2014), and Chen et al. (2021a) study this problem in a stochastic setting with i.i.d. demand arrivals. In these works, the authors use the primal-dual technique to design algorithms with sub-linear regret, where the algorithms aim to learn the optimal dual variables associated with resource constraints. The primal-dual technique is also effective for adversarial demand arrivals even though attaining sub-linear regret in adversarial settings is generally impossible (Mehta et al. 2007, Buchbinder et al. 2007, Golrezaei et al. 2014). See also Balseiro et al. (2020) for a recent work that shows how the primal-dual technique results in well-performing algorithms for various demand processes. While in our work, we do not rely on the primal-dual technique, our work contributes to this literature by presenting a model that—with the help of the sample information—bridges the gap between stochastic and adversarial arrivals, allowing us to bypass the impossibility results in the adversarial settings which is obtained by Ball and Queyranne (2009).

## 2. Model

We consider a decision-maker who would like to allocate (identical) units of a resource/service to two types of unit-demand agents. (In Section 7.4, we discuss how to

extend our algorithm (Algorithm 1) to a setting with more than two types.) The expected reward of agents of type  $i \in \{1, 2\}$  upon receiving one unit of the resource is  $r_i \in (0, 1)$ , where without loss of generality, we assume that  $r_1 > r_2$  and define  $\alpha = \frac{r_2}{r_1} \in (0, 1)$ . The decision-maker, however, is not aware of the expected reward of agents; she does not even know which type of agents attains a higher expected reward. As stated in the introduction, this setting captures scenarios where we would like to allocate new drugs, services, and products to customers. Given the lack of knowledge about the expected rewards, before the allocation period starts, the decision-maker aims to collect some information about the unknown expected rewards of agents during a *test period*.

### 2.1. Test Period and Sample Information

During the test period, the decision-maker aims to outreach the market by, for example, making a public announcement. Let  $h, \ell \geq 0$  be respectively the market size of agents of types 1 and 2, i.e., the total number of agents of types 1 and 2 that are interested in the resource. The market size  $h$  and  $\ell$ , which are unknown to the decision-maker, can take any arbitrary values; that is, they are chosen adversarially. We assume that during the test period, with probability  $p \in (0, 1)$ , each of  $h + \ell$  agents sees and responds to the outreach program run by the decision-maker, where the *sampling probability*  $p$  is known to the decision-maker. (In Section 7.3, we show that our algorithm and results can be simply extended to a setting where each agent of type  $i \in \{1, 2\}$  reacts to the outreach program with probability  $p_i$ , where  $p_1 \neq p_2$ .)

Note that we do not require  $p$  to be a constant. Our results not only hold for any constant value of the sampling probability  $p$ , but also hold when  $p = \omega(\frac{1}{\sqrt{m}})$ , where  $m$  is the total number of resources when the test period ends. Furthermore, the assumption of knowing the sampling probability is motivated by the fact that the outreach program is designed by the decision-maker herself and hence an accurate estimate of the sampling probability  $p$  is available to her. Nonetheless, in Section 3, we show that without knowing  $p$ , achieving asymptotic optimality is not possible. See also our case study in Section 6 and our numerical studies in Appendix K where we investigate the robustness of our proposed algorithm to the lack of exact knowledge about  $p$ .

Any agent who responds to the outreach program gets one unit of the resource and reveals his realized reward to the decision-maker.<sup>8</sup> Here, for simplicity, we assume that the realized reward of type  $i \in \{1, 2\}$  agents is either 0 or 1. This assumption can be relaxed by letting the realized rewards take more than two values. Let  $s_i$ ,  $i \in \{1, 2\}$ , be the (random) number of agents of type  $i$  that are reached/*sampled* during the test period. Note that  $s_1$  is drawn from a binomial distribution with parameters  $h$  and  $p$  (i.e.,  $s_1 \sim \text{Bin}(h, p)$ ) and  $s_2 \sim \text{Bin}(\ell, p)$ . The set of realized rewards of  $s_i$  sampled agents is denoted by  $\boldsymbol{\rho}_i \in \{0, 1\}^{s_i}$ , where  $\boldsymbol{\rho}_i = \{\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,s_i}\}$  and  $\rho_{i,j}$ ,  $j \in [s_i]$ , is drawn from a Bernoulli distribution with the success probability of  $r_i$ . Throughout the paper, we denote  $(s_1, s_2)$  and  $(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$  with  $\mathbf{s}$  and  $\boldsymbol{\rho}$ , respectively. In addition, we refer to  $\boldsymbol{\psi} := (\mathbf{s}, \boldsymbol{\rho})$  as the *sample information* that decision-maker obtains during the test period. (Note that  $\boldsymbol{\rho}$  depends on the random variables  $\mathbf{s}$ . But, to simplify the exposition, we do not show the dependency of  $\boldsymbol{\rho}$  to  $\mathbf{s}$ .)

## 2.2. Online Allocation Period

Having described the test period, we are now ready to explain the allocation period. Let  $m \geq 2$  be the number of available resources at the beginning of the allocation period. During this period, the rest of the market, i.e.,  $h - s_1$  type 1 agents and  $\ell - s_2$  type 2 agents arrive one by one over time in an arbitrary order. We denote the number of agents of type  $i$  during the allocation period with  $n_i$  and we further denote  $(n_1, n_2)$  with  $\mathbf{n}$ . Let  $z_t \in \{1, 2\}$ ,  $t = 1, 2, \dots$ , be the type of the agent in time period  $t$  within the allocation period. The agent type  $z_t$ , which can represent any available information at the time of the decision, is observable to the decision-maker at time period  $t$ , but given the online nature of the problem during the allocation period,  $z_{t'}$  for any  $t' > t$  is not observable at time period  $t$ . We define  $I = (z_t)_{t \geq 1}$  as the *online arrival sequence* and note that  $n_i = |\{t \leq |I| : z_t = i\}|$ ,  $i \in \{1, 2\}$ .

Upon the arrival of the agent of type  $z_t$  in time period  $t$ , the decision-maker has to make an irrevocable acceptance/rejection decision regarding that agent. If the decision-maker accepts the agent, she allocates the agent one unit of the resource. Otherwise, no resource will be allocated to the agent, and that agent will not come back. If the agent gets accepted, he reveals his realized reward to the decision-maker.

<sup>8</sup>The assumption that any agent in the test period gets a resource can be relaxed as we show in Section 7.1. More specifically, there we show that our asymptotic result still holds when the number of resources in the test period is in the order of  $\omega(\sqrt{m})$ , which includes the case of  $\Theta(m)$ .

In this work, to ease the analysis, we will focus on the continuous version of our problem, where the resources are divisible and we are allowed to partially accept agents. We can still capture all the core ideas regarding sample information in continuous models. For the discrete model, our algorithm can still be applied and are asymptotically optimal as  $m \rightarrow \infty$ . To deal with small values of  $m$ , one can apply a technique used in Ball and Queyranne (2009) to randomly accept or reject agents, obtaining the same performance guarantees in expectation.

### 2.3. Performance Measure

The goal of the decision-maker during the allocation period is to yield high total rewards while being uncertain about the number and the order of agents, as well as, their expected rewards. For an algorithm  $A$ , online arrival sequence  $I$ , and the sample information,  $\psi = (\mathbf{s}, \boldsymbol{\rho})$ , let  $\text{REW}_A(I, \psi)$  be the cumulative expected reward of algorithm  $A$  across all the time periods. We measure the performance of an algorithm  $A$  using the following competitive ratio (CR) definition, which compares our algorithm to the optimal clairvoyant benchmark that knows the arrival sequence  $I$  and the expected reward of agents ( $r_i$ 's) in advance.<sup>9</sup> Let  $\text{OPT}(I)$  be the optimal clairvoyant cumulative expected reward that can be obtained from  $I$  using  $m$  units of the resource. The CR of an algorithm  $A$  is then defined as

$$\text{CR}_A = \inf_{(h, \ell)} \mathbb{E}_{\psi} \left[ \inf_I \frac{\mathbb{E}[\text{REW}_A(I, \psi)]}{\text{OPT}(I)} \right]. \quad (1)$$

Here, the inner expectation is with respect to (w.r.t.) any randomness in algorithm  $A$ , and the outer expectation is w.r.t. the arrival sequence  $I$  and the sample information  $\psi$ . Note that given  $(h, \ell)$ , the arrival sequence  $I$  and more precisely, the number of agents of type  $i \in \{1, 2\}$  in  $I$  (denoted by  $n_i$ ), is random. Further, observe that in our definition of CR, we take infimum over (i) the size of the market, i.e.,  $h, \ell$ , and (ii) the order of arrivals in  $I$ .

## 3. Impossibility Results

In this section, to shed light on the challenges in our setting and the necessity of some of the assumptions we made, we present alternative scenarios for which we show that it is impossible to obtain a CR that goes to one as the number of resources  $m$  goes to infinity. These scenarios are listed below.

<sup>9</sup> Observe that the optimal clairvoyant benchmark does not know the realized rewards in advance. This optimal clairvoyant benchmark is consistent with that used in Ball and Queyranne (2009). Nevertheless, in Section 7.2, we discuss a stronger benchmark that knows the realized rewards. We show that under a strong clairvoyant benchmark that knows the realized reward of agents, it is not possible to design an algorithm whose competitive ratio w.r.t. this strong benchmark goes to one as  $m$  goes to infinity.

- *Scenario 1: No Sample Information.* Consider the setting in Section 2 where the rewards  $r_i$  are unknown and the arrival sequence during the allocation period is chosen adversarially. But, unlike our original setting in Section 2, the sampling probability  $p = 0$ , and hence, no sample information is available. For this scenario, we show that the CR is upper bounded by  $1/(2 - \alpha)$ , where  $\alpha = r_2/r_1$  and  $1/(2 - \alpha)$  is the bound in Ball and Queyranne (2009). This impossibility result highlights the importance of the sample information in achieving the asymptotically optimal CR. See also Section 5.2 where we show that even with  $p = O(1/m)$ , the CR cannot exceed  $1/(2 - \alpha)$ .
- *Scenario 2: Unknown Sampling Probability.* Consider the setting in Section 2 with sample information, unknown rewards, and adversarial arrival sequences. But, assume that the sampling probability  $p$  is completely unknown to the decision-maker. For this scenario, we show an upper bound of  $1/(2 - \alpha)$  on the CR. This impossibility result emphasizes the importance of knowing the sampling probability. Note that as stated earlier, similar to our setting, Correa et al. (2021) also assume that the sampling probability  $p$  is known and they crucially use the knowledge of  $p$  to design optimal algorithms for the secretary problem.

**PROPOSITION 1 (Impossibility Results).** *For the scenarios 1 and 2, defined above, the CR of any deterministic and randomized algorithm is upper bounded by  $1/(2 - \alpha)$ , where  $\alpha = r_2/r_1$ .*

The proof of Proposition 1 is presented in Section A in the appendix.

#### 4. Online Resource Allocation Algorithm with Samples

In light of our impossibility results, here, under our original model presented in Section 2, we present an online resource allocation algorithm that is asymptotically optimal as the number of resources  $m$  goes to infinity; that is, the CR of our algorithm converges to one as  $m$  goes to infinity.

In our setting, the decision-maker receives some information about the number of agents of each type in the online arrival sequence. The decision-maker further obtains some partial information about the expected reward of each type. Recall that the decision-maker observes  $s_i$  samples,  $i \in \{1, 2\}$ , for each type  $i$  agents where  $s_1 \sim \text{Bin}(h, p)$  and  $s_2 \sim \text{Bin}(\ell, p)$ . In addition, the decision-maker observes the realized reward of  $s_i$  agents of type  $i \in \{1, 2\}$ , denoted by  $\boldsymbol{\rho}_i = (\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,s_i})$ .

Our proposed algorithm (Algorithm 1) takes advantage of both  $\mathbf{s} = (s_1, s_2)$  and  $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ . The algorithm uses these pieces of information to obtain an estimate of the expected reward of each type  $i$ , denoted by  $\widehat{r}_i$ . In particular, if  $s_i > 0$ , the algorithm simply uses the sample average of the realized reward observed in  $\boldsymbol{\rho}_i$  as an estimate for  $r_i$ ; that is,

$$\widehat{r}_i = \frac{\sum_{j=1}^{s_i} \rho_{i,j}}{s_i} \quad i \in \{1, 2\}.$$

When  $s_i = 0$ ,  $\widehat{r}_i$  is randomly drawn from a uniform distribution in  $[0, 1]$ .

Having access to these estimates, the algorithm follows a protection level policy to protect the type that has the highest estimated expected reward. The definition of a protection level policy for type  $i \in \{1, 2\}$  and protection level  $x$  is stated below in Definition 1.

**DEFINITION 1 (PROTECTION LEVEL POLICY FOR TYPE  $i$  AND PROTECTION LEVEL  $x$ ).** In this algorithm, agents of type  $i$  are always accepted unless there is no resource left. Agents of type  $-i$  will be accepted if the number of accepted type  $-i$  agents is less than  $m - x$  and there is resource left.<sup>10</sup>

When  $\widehat{r}_1 > \widehat{r}_2$ , then the algorithm assigns a protection level of  $x_1 = \min\{m, s_1 \frac{1-p}{p}\}$  for type 1 agents, where we note that conditioned on  $s_1$ ,  $s_1 \frac{1-p}{p}$  is equal to the expected number of type 1 agents in the online arrival sequence, i.e.,  $n_1$ . Recall that  $n_1 = h - s_1$  and  $s_1 \sim \text{Bin}(h, p)$ . On the other hand, when  $\widehat{r}_2 \geq \widehat{r}_1$ , the algorithm assigns a protection level of  $x_2 = \min\{m, s_2 \frac{1-p}{p}\}$  for type 2 agents. The description of the algorithm can be found in Algorithm 1.

---

### Algorithm 1 Online Resource Allocation Algorithm with Samples

---

**Input:** The number of resources  $m$  and sample information  $\boldsymbol{\psi} = (\mathbf{s}, \boldsymbol{\rho})$ , where  $\mathbf{s} = (s_1, s_2)$ ,  $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ , and for any  $i \in \{1, 2\}$ ,  $\boldsymbol{\rho}_i = \{\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,s_i}\}$ .

1. If the number of samples for type  $i \in \{1, 2\}$ , i.e.,  $s_i$ , is positive, define

$$\widehat{r}_i = \frac{\sum_{j=1}^{s_i} \rho_{i,j}}{s_i}$$

as an estimate of the expected reward of type  $i \in \{1, 2\}$ . Otherwise,  $\widehat{r}_i \sim \text{uniform}(0, 1)$ , where  $\text{uniform}(0, 1)$  is the uniform distribution in  $[0, 1]$ .

2. If  $\widehat{r}_1 > \widehat{r}_2$ , set the protection level  $x_1 = \min\{m, s_1 \frac{1-p}{p}\}$  for type 1, and run the protection level policy for type 1 agents and protection level  $x_1$  as per Definition 1.
  3. If  $\widehat{r}_1 \leq \widehat{r}_2$ , set the protection level  $x_2 = \min\{m, s_2 \frac{1-p}{p}\}$  for type 2 and run the protection level policy for type 2 agents and protection level  $x_2$  as per Definition 1.
- 

<sup>10</sup>In this algorithm, the last accepted type  $j \in \{1, 2\}$  agent may get accepted partially.

We highlight that in the described algorithm, the adversary’s choice of  $h$  and  $\ell$  influences the algorithm’s estimate for the expected rewards, as well as, the protection levels  $x_1$  and  $x_2$ . Furthermore, the adversary’s choices impact the algorithm’s decision about protecting a certain type. Note that when the estimated expected rewards  $\hat{r}_i$ ,  $i \in \{1, 2\}$ , are noisy, the algorithm may end up protecting a wrong type (i.e., the type with a lower reward). Despite these challenges, as we show in Theorem 1 and Proposition 2, Algorithm 1 manages to perform very well. Theorem 1 presents a lower bound on the CR of Algorithm 1 for any value of  $m$  and  $p$ , where we recall that  $m$  is the number of resources and  $p$  is the sampling probability. Further, Proposition 2 shows how this lower bound scales with  $m$  and  $p$  as  $m$  goes to infinity. In particular, Proposition 2 shows that the asymptotic CR of Algorithm 1 is  $1 - \Theta(1/(p\sqrt{m}))$ .

Later in Section 5, we show this bound is tight in the sense that no randomized or deterministic algorithm can break the asymptotic bound of  $1 - \Theta(1/(p\sqrt{m}))$ . The asymptotic optimality of Algorithm 1 is quite remarkable given the simplicity of the algorithm and our strong adversary. Note that the adversary through choosing the market size  $h$  and  $\ell$  can control to what extent the sample information is informative. When  $h$  and  $\ell$  are both large, the number of agents in the sample is also large and hence the sample information can be very useful in reducing uncertainty in demand and expected rewards of agents. However, when either  $h$  or  $\ell$  is small—even when the number of resources  $m$  goes to infinity and the sampling probability is constant—the sample information may not be very informative, as it does not even allow the decision-maker to obtain an accurate estimate of rewards. As a result, our algorithm may end up protecting the wrong type. Thus, the fact that our simple algorithm obtains an asymptotically tight CR of  $1 - \Theta(1/(p\sqrt{m}))$  is outstanding.

In addition, Algorithm 1 does not use the realization of the reward of arriving agents during the allocation period. One may wonder if using such information would help us design a better algorithm. The answer is no. Our upper bound result in Section 5 shows that adaptive algorithms that further use the realized rewards of agents during the allocation period cannot obtain a better asymptotic CR. Intuitively speaking, this is because the arrival sequence is chosen by an adversary who can manipulate how fast the rewards are learned by choosing the order of arrivals. That is, the adversary can prevent the decision-maker to have a good estimate about the rewards when having such estimates is crucial.

Therefore, using the realized rewards during the allocation period does not give algorithms an edge.

Nonetheless, one can consider an adaptive version of Algorithm 1. In this algorithm, estimates of rewards  $\hat{r}_i$  are updated once an agent is accepted: we can again let  $\hat{r}_i$  be the empirical mean of realized rewards of all agents of type  $i$  who receives the resource. At any time  $t$ , if we observe  $\hat{r}_i > \hat{r}_j$ , we give type  $i$  agent protection level  $x_i = \min\{m - Z_{j,t}, s_i \frac{1-p}{p} - Y_{i,t}\}$ , where  $Y_{i,t}$  is the number of type  $i$  agents arrived before time  $t$  and  $Z_{j,t}$  is the number of type  $j$  agents accepted before time  $t$ . In Section 6, via our case study, we evaluate this adaptive algorithm, and we show that even in realistic arrival scenarios, the CR of this adaptive algorithm is almost the same as that of Algorithm 1, verifying our theoretical result.

#### 4.1. Competitive Ratio of Algorithm 1

Here, we present a lower bound on the CR of Algorithm 1. To do so, we start with the following lemma. This lemma shows that to characterize a lower bound on the CR of the algorithm, it suffices to only consider ordered arrival sequences in which any type 2 agent arrives before any type 1 agent. The proofs of all the lemmas in this section are presented in Section B in the appendix.

**LEMMA 1 (Worst Order).** *For any realization  $n_1, n_2$ , where  $n_i$  is the number of type  $i$  agents in the arrival sequence, let  $I_{order}$  be an ordered arrival sequence under which  $n_2$  type 2 agents arrived first, followed by  $n_1$  type 1 agents. Let  $I$  be any arrival sequence that contains  $n_2$  type 2 agents and  $n_1$  type 1 agents. Then, under Algorithm 1 (denoted by  $A$ ), we have*

$$\frac{\mathbb{E}[\text{REW}_A(I_{order}, \boldsymbol{\psi})]}{\text{OPT}(I_{order})} \leq \frac{\mathbb{E}[\text{REW}_A(I, \boldsymbol{\psi})]}{\text{OPT}(I)}. \quad (2)$$

With a slight abuse of notation, let  $\text{OPT}(\mathbf{n})$  be the optimal clairvoyant cumulative expected reward under an online ordered arrival sequence  $I$  with  $n_i \in \{1, 2\}$  type  $i$  agents. Similarly,  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})$  is the (expected) reward of Algorithm 1 under  $\boldsymbol{\rho}$  and an online ordered arrival sequence  $I$  with  $n_i$  type  $i$  agents. Then, Lemma 1 allows us to rewrite the CR of Algorithm 1 as follows

$$\begin{aligned} \text{CR}_A &= \inf_{(h,\ell)} \mathbb{E}_{\boldsymbol{\psi}} \left[ \inf_I \frac{\mathbb{E}[\text{REW}_A(I, \boldsymbol{\psi})]}{\text{OPT}(I)} \right] = \inf_{(h,\ell)} \mathbb{E}_{\boldsymbol{\psi}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] \\ &= \inf_{(h,\ell)} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right]. \end{aligned} \quad (3)$$



Note that in the inner expectation of the last expression, we take expectation w.r.t. any randomness in the algorithm and  $\boldsymbol{\rho}$  given a realization of  $\mathbf{n}$ .

To bound  $\text{CR}_A$ , we define the following *good event*, denoted by,  $\mathcal{G}(\boldsymbol{\psi}; (h, \ell))$ , as the event under which Algorithm 1 correctly assumes that type 1 agents have a higher expected reward than type 2 agents when the sample information is  $\boldsymbol{\psi}$  and the market size for type 1 and 2 agents is respectively  $h$  and  $\ell$ :

$$\mathcal{G}(\boldsymbol{\psi}; (h, \ell)) = \{\widehat{r}_1 > \widehat{r}_2\} = \left\{ \frac{\sum_{j=1}^{s_1} \rho_{1,j}}{s_1} > \frac{\sum_{j=1}^{s_2} \rho_{2,j}}{s_2} \right\}. \quad (4)$$

Then, given a realization of  $\mathbf{s}$  or equivalently  $\mathbf{n}$ , we have

$$\begin{aligned} \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})] &= \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) \cdot \mathbf{1}(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))] + \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) \cdot \mathbf{1}(\mathcal{G}^c(\boldsymbol{\psi}; (h, \ell)))] \\ &= \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) \mid \mathcal{G}(\boldsymbol{\psi}; (h, \ell))] \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell))) \\ &\quad + \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) \mid \mathcal{G}^c(\boldsymbol{\psi}; (h, \ell))] \cdot \Pr(\mathcal{G}^c(\boldsymbol{\psi}; (h, \ell))), \end{aligned} \quad (5)$$

where  $\mathbf{1}(\mathcal{A})$  is 1 if an event  $\mathcal{A}$  occurs and zero otherwise. Here, given  $\mathbf{n} = (n_1, n_2)$ , the expectation is taken on  $\rho_{1,j} \sim \text{Ber}(r_1)$  for  $j \in [s_1]$  and  $\rho_{2,j} \sim \text{Ber}(r_2)$  for  $j \in [s_2]$ , where  $\text{Ber}(r)$  is a Bernoulli distribution with a success probability of  $r \in [0, 1]$ . Further,  $\mathcal{G}(\boldsymbol{\psi}; (h, \ell))$  is the good event defined in Equation (4). In Equation (5), the expression  $\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) \mid \mathcal{G}(\boldsymbol{\psi}; (h, \ell))]$  presents the reward of Algorithm 1 under a specific realization of  $\mathbf{n}$  when the good event  $\mathcal{G}(\boldsymbol{\psi}; (h, \ell))$  happens. Note that for a given realization of  $\mathbf{n}$  and under the good event, the algorithm's action and hence its reward does not depend on  $\boldsymbol{\rho}$ . This is because when the good event happens, the algorithm assigns a protection level  $x_1$  to type 1, where  $x_1$  only depends on  $s_1$ , not  $\boldsymbol{\rho}$ . Finally, in  $\Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))$ , we take expectation with respect to  $\boldsymbol{\rho}$  and any randomness in the algorithm in defining  $\widehat{r}_i$ ,  $i \in [2]$ , given a realization of  $\mathbf{n}$ .

To bound the CR of Algorithm 1, we consider the following three cases based on the number of type 1 and 2 agents chosen by the adversary (i.e.,  $h$  and  $\ell$ ):

*Case 1.* In this case, the number of type 1 agents is small (less than  $\sqrt{m}$ ). That is,  $(h, \ell) \in \mathcal{R}_1$ , where  $\mathcal{R}_1 = \{(h, \ell) : h < \sqrt{m}, \ell \geq 0\}$ .

*Case 2.* In this case, while the number of type 1 agents is large, the number of type 2 agents is small. That is,  $(h, \ell) \in \mathcal{R}_2$ , where  $\mathcal{R}_2 = \{(h, \ell) : h \geq \sqrt{m}, \ell < \sqrt{m}\}$ .

*Case 3.* In this case, the numbers of both type 1 and 2 agents are large. That is,  $(h, \ell) \in \mathcal{R}_3$ , where  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ .

Observe that the worst case CR of Algorithm 1 can be written as

$$\min_{i \in [3]} \left\{ \inf_{(h, \ell) \in \mathcal{R}_i} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] \right\}, \quad (6)$$

where by Equation (5), for any  $i \in [3]$ , we have

$$\begin{aligned} \inf_{(h, \ell) \in \mathcal{R}_i} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] &= \inf_{(h, \ell) \in \mathcal{R}_i} \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) | \mathcal{G}(\boldsymbol{\psi}; (h, \ell))] \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) | \mathcal{G}^c(\boldsymbol{\psi}; (h, \ell))] \cdot \Pr(\mathcal{G}^c(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right\}. \end{aligned}$$

In the rest of the paper, we use the following shorthand notation to simplify the exposition:

$$\begin{aligned} \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) | \mathcal{G}(\boldsymbol{\psi}; (h, \ell))] &= \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \\ \mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}) | \mathcal{G}^c(\boldsymbol{\psi}; (h, \ell))] &= \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c). \end{aligned}$$

With this notation, we have

$$\begin{aligned} \inf_{(h, \ell) \in \mathcal{R}_i} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] &= \inf_{(h, \ell) \in \mathcal{R}_i} \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c) \cdot \Pr(\mathcal{G}^c(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right\}. \quad (7) \end{aligned}$$

Having defined this shorthand notation, in the following, we present three main lemmas: Lemmas 2, 3, and 4, where each of these lemmas provides a lower bound on  $\inf_{(h, \ell) \in \mathcal{R}_i} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right]$  for one of the regions  $\mathcal{R}_i$ ,  $i \in [3]$ .

**LEMMA 2 (Region  $\mathcal{R}_1$ ).** *Let  $\mathcal{R}_1 = \{(h, \ell) : h < \sqrt{m}, \ell \geq 0\}$ . Then, we have*

$$\inf_{(h, \ell) \in \mathcal{R}_1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] \geq CR_1,$$

where

$$CR_1 = \min \left\{ \left( 1 - \frac{1-p}{p\sqrt{m}} \right)^+, 1 - \frac{r_1 - r_2}{(r_1 - r_2) + r_2\sqrt{m}} \right\},$$

and  $y^+ = \max\{y, 0\}$ .

Lemma 2 shows when  $(h, \ell)$  falls into region  $\mathcal{R}_1$ , the CR of Algorithm 1 is lower bounded by  $CR_1$ . From the expression of  $CR_1$ , the asymptotic bound in region  $\mathcal{R}_1$  is  $1 - \Theta(1/(p\sqrt{m}))$ , and based on the result we will present in Section 5, this bound is tight with respect to

both  $p$  and  $m$ . The bound of  $1 - \Theta(1/(p\sqrt{m}))$  is achieved when the total number of arrivals is less than  $m$  (i.e.,  $n_1 + n_2 \leq m$ ). When the total number of arrivals is greater than  $m$ , a better asymptotic bound of  $1 - \Theta(1/\sqrt{m})$ —which does not depend on  $p$ —can be obtained.

To show Lemma 2, by Equation (7), we need to bound  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G}) \cdot \Pr(\mathcal{G}(\psi; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right]$  and  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G}^C) \cdot \Pr(\mathcal{G}^C(\psi; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right]$  respectively. As in this region,  $h$  can be very small, we cannot use any concentration inequality to bound the expectation over  $\mathbf{n}$ . Instead, we bound  $\frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G})}{\text{OPT}(\mathbf{n})}$  and  $\frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})}$  for any realization of  $\mathbf{n}$  by considering different ranges for the number of agents of type  $i \in [2]$  in the online arrival sequence.

**LEMMA 3 (Region  $\mathcal{R}_2$ ).** *Let  $\mathcal{R}_2 = \{(h, \ell) : h \geq \sqrt{m}, \ell < \sqrt{m}\}$ . Then, we have*

$$\inf_{(h, \ell) \in \mathcal{R}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \rho)]}{\text{OPT}(\mathbf{n})} \right] \geq CR_2,$$

where

$$CR_2 = \min \left\{ \left(1 - \frac{1}{\sqrt{m}}\right) \left(1 - \frac{1}{m}\right), \frac{h_1}{m} \right\}.$$

Here,  $h_1 = h_0(1 - p) - \sqrt{p(1 - p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1 - p} \geq yp + \sqrt{y}\}$ , and  $\beta = 0.4215 \cdot \frac{p^2 + (1 - p)^2}{p(1 - p)}$ .

Lemma 3 shows when  $(h, \ell)$  falls into region  $\mathcal{R}_2$ , the CR of Algorithm 1 is lower bounded by  $CR_2$ . In the expression of  $CR_2$ ,  $h_0 = \frac{m - \sqrt{m}}{1 - p} + \frac{1}{2p^2} + \sqrt{\frac{1}{4p^4} + \frac{(m - \sqrt{m})}{p^2(1 - p)}}$ , and as  $p = \omega(1/\sqrt{m})$ , we have  $h_0 = \frac{m}{1 - p} - \Theta(\frac{\sqrt{m}}{1 - p})$ ,  $h_1 = m - \Theta(\sqrt{m})$ , and  $\frac{h_1}{m} = 1 - \Theta(1/\sqrt{m})$ . This implies that the asymptotic bound in region  $\mathcal{R}_2$  is  $1 - \Theta(1/\sqrt{m})$ , which is tight with respect to  $m$ , but is not related to  $p$ .

To show Lemma 3, we consider three factors which contribute to the loss of Algorithm 1, where for a realization of  $\mathbf{n}$  and  $\rho$ , the (normalized) loss of Algorithm  $A$  is  $\frac{\text{OPT}(\mathbf{n}) - \text{REW}_A(\mathbf{n}, \rho)}{\text{OPT}(\mathbf{n})} = 1 - \frac{\text{REW}_A(\mathbf{n}, \rho)}{\text{OPT}(\mathbf{n})}$ . These factors are: (i) protecting the wrong (low-reward) type, (ii) over-protecting the right (high-reward) type, and (iii) under-protecting the right type. In case (i), as the number of type 2 agents is very small ( $\ell < \sqrt{m}$ ), the loss caused by wrongfully protecting type 2 agents is bounded by  $1/\sqrt{m}$ , and hence, the CR is lower bounded by  $1 - 1/\sqrt{m}$ . This bound is obtained by considering the worst case among any realization of  $n_1$  and  $n_2$ .

To bound the expected loss caused by over- or under-protecting type 1 agents for cases (ii) and (iii), we use concentration inequalities, as the number of type 1 agents is large under

region  $\mathcal{R}_2$ . In doing so, we face one main challenge: In computing the expected (normalized) loss (or equivalently the expected CR), it is hard to compute the expectation w.r.t.  $\mathbf{n}$  as both the reward/loss of the algorithm (the numerator in the CR, i.e.,  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})$ ) and the optimal clairvoyant solution (the denominator in the CR, i.e.,  $\text{OPT}(\mathbf{n})$ ) are random variables. Recall that after choosing  $h$  and  $\ell$ , the number of agents  $n_i$ 's in the arrival sequence are random.

To overcome this challenge, based on the value of  $h$ , we split the region  $\mathcal{R}_2$  into two sub-regions  $\underline{\mathcal{R}}_2$  and  $\overline{\mathcal{R}}_2$ , where  $\underline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h < h_0\}$  and  $\overline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h \geq h_0\}$ . In each of the sub-regions, we can deal with the aforementioned expectation challenge by bounding the optimal clairvoyant solution in the denominator. The splitting allows us to (i) provide a tight bound on  $\text{OPT}(\mathbf{n})$  in each of the sub-regions. This then further enables us to only focus on computing the expectation for the numerator. We show that for the sub-region  $\underline{\mathcal{R}}_2$ , the CR can be lower bounded by  $(1 - \frac{1}{\sqrt{m}})(1 - \frac{1}{m})$ , and for the sub-region  $\overline{\mathcal{R}}_2$ , the CR can be bounded by  $\min\{1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m}\}$ . As the former term (i.e.,  $(1 - \frac{1}{\sqrt{m}})(1 - \frac{1}{m})$ ) is dominated by the latter (i.e.,  $\min\{1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m}\}$ ), we have the CR for  $\mathcal{R}_2$  is at least  $\min\left\{\left(1 - \frac{1}{\sqrt{m}}\right)\left(1 - \frac{1}{m}\right), \frac{h_1}{m}\right\}$ .

**LEMMA 4 (Case 3: region  $\mathcal{R}_3$ ).** *Let  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ . Then,  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right]$  is at least*

$$\begin{cases} \overline{CR}_3, & m \geq m_1, \\ \underline{CR}_3, & m < m_1. \end{cases},$$

where

$$\begin{aligned} \overline{CR}_3 &= \left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{m}p - m^{1/4})^2}\right)^2 W \\ \underline{CR}_3 &= V \cdot \min \left\{ \left(1 - \frac{1}{m^2}\right)W, \frac{1}{2} \left(1 - \frac{1}{m^2}\right)W + \frac{1}{2} \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right)\alpha, 1 - \frac{1-p}{pm} \ell_1 \right\} \right\}. \end{aligned}$$

Here,  $V = 1 - 2(1-p)\sqrt{m}$ ,  $W = \min\left\{1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m}\right\}$ ,  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ ,  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ ,  $m_1 = \min_{y \geq 1/p^4} \{y : r_1 - r_2 > \frac{2}{y^{1/8} \sqrt{y^{1/4} p - 1}}\}$ ,  $\ell_1 = \sqrt{p(1-p)\ell_0} + \frac{\beta}{\sqrt{m}} + \ell_0 p$ , and  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$ .

**REMARK 1.** We provide the exact formula for  $\ell_0$  and  $m_1$  here. Note that  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$  can be written as  $\frac{m}{1-p} + \frac{1 + \sqrt{4m(1-p)+1}}{2(1-p^2)}$ . In addition,  $m_1 = \min_{y \geq 1/p^4} \{y :$

$r_1 - r_2 > \frac{2}{y^{1/8}\sqrt{y^{1/4}p-1}}$  can be written as  $m_1 = \left(\frac{b+\sqrt{b^2+4ac}}{2a}\right)^4$ , where  $a = (4 - 3p)p$ ,  $b = (1 - p)(4 - 3p)^{\frac{1}{2}}$  and  $c = \frac{4(1-p)^2}{(r_1-r_2)^2}$ . Further note that as  $m \rightarrow \infty$ ,  $\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REWARD}(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] = \overline{\text{CR}}_3 = 1 - \max\{\Theta(1/\sqrt{m}), \Theta(1/(p^2m))\}$ .

Lemma 4 shows that when  $(h, \ell)$  falls into region  $\mathcal{R}_3$ , the CR of Algorithm 1 is lower bounded by  $\overline{\text{CR}}_3$ . From the expression of  $\overline{\text{CR}}_3$  and Remark 1, the asymptotic bound in region  $\mathcal{R}_3$  is  $1 - \max\{\Theta(1/\sqrt{m}), \Theta(1/(p^2m))\}$ . This shows that our algorithm has a better asymptotic CR in region  $\mathcal{R}_3$ , compared with region  $\mathcal{R}_1$ . (Recall that the asymptotic CR in region  $\mathcal{R}_1$  is  $1 - \Theta(1/p\sqrt{m})$  and  $p = \omega(1/\sqrt{m})$ .) The fact that region  $\mathcal{R}_1$  is more challenging is because under region  $\mathcal{R}_1$ , the number of type 1 agents, i.e.,  $h$ , is small and hence the sample information may not be informative enough.

To show Lemma 4, the main idea is similar to the one in region  $\mathcal{R}_2$ . As the number of type 1 agents is still large in  $\mathcal{R}_3$ , to bound the loss caused by over- or under-protecting type 1 agents, we can use concentration inequalities to show that either the bad situation (i.e., over- or under-protecting type 1 agents) does not happen or if it happens, it does not have a significant amount of loss.

However, as the number of type 2 agents is also large in  $\mathcal{R}_3$ , to bound the loss caused by wrongfully protecting type 2 agents, we need to further split the analysis into two parts based on the number of resources  $m$ . When  $m \geq m_1$ , we show that the probability that Algorithm 1 wrongfully protects type 2 agents is small because, in this case, we have enough samples to get an accurate estimation of the reward of each type. When  $m < m_1$ , we directly compute this loss. The main challenge is still dealing with the expectation of a fraction of two random variables. To handle this challenge, we again split region  $\mathcal{R}_3$  into two sub-regions:  $\underline{\mathcal{R}}_3$  and  $\overline{\mathcal{R}}_3$ , where  $\overline{\mathcal{R}}_3 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell > \ell_0\}$  and  $\underline{\mathcal{R}}_3 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell \leq \ell_0\}$ . By providing a proper bound for the optimal clairvoyant solution in the denominator, we then show that the CR under region  $\underline{\mathcal{R}}_3$  is at least  $\underline{\text{CR}}_3$  and the CR under region  $\overline{\mathcal{R}}_3$  is at least  $\overline{\text{CR}}_3$ .

Next, by putting the results of Lemmas 2, 3, and 4, we get the following theorem.

**THEOREM 1 (Competitive ratio of Algorithm 1).** *Consider the model presented in Section 2, where the expected rewards of  $r_i$ ,  $i \in \{1, 2\}$  are unknown to the decision-maker. Let  $h_0 = \min\{y \geq 0 : \frac{p(m-\sqrt{m})}{1-p} \geq yp + \sqrt{y}\} = \Theta(\frac{m}{1-p})$ ,  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}} = 1 - \Theta(\frac{(1+\sqrt{p})}{\sqrt{m}})$ ,  $V = 1 - 2(1-p)^{\sqrt{m}} = 1 - \Theta((1-p)^{\sqrt{m}})$ ,  $W = \min\left\{1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m}\right\} = 1 - \Theta(\frac{1}{\sqrt{m}})$ ,*

$\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\} = \Theta(\frac{m}{1-p})$ ,  $\ell_1 = \sqrt{p(1-p)\ell_0} + \frac{\beta}{\sqrt{m}} + \ell_0 p = 1 - \Theta(\frac{(1+\sqrt{p})}{\sqrt{m}})$ ,  $\alpha = \frac{r_2}{r_1} > 0$ , and  $\beta = 0.4215 \cdot \frac{p^2+(1-p)^2}{p(1-p)}$ . Then, for any  $m \geq 2$  and  $p \in (0, 1)$ , the CR of Algorithm 1, denoted by  $CR_A$ , is lower bounded by:

$$CR_A \geq \begin{cases} \min\{CR_1, CR_2, \overline{CR}_3\}, & m \geq m_1, \\ \min\{CR_1, CR_2, \underline{CR}_3\}, & m < m_1. \end{cases},$$

where  $m_1 = \min_{y \geq 1/p^4} \{y : r_1 - r_2 > \frac{2}{y^{1/8}\sqrt{y^{1/4}p-1}}\}$ , and

$$\begin{aligned} CR_1 &= \min \left\{ \left(1 - \frac{1-p}{p\sqrt{m}}\right)^+, 1 - \frac{r_1 - r_2}{(r_1 - r_2) + r_2\sqrt{m}} \right\}, \\ CR_2 &= \min \left\{ \left(1 - \frac{1}{\sqrt{m}}\right)\left(1 - \frac{1}{m}\right), \frac{h_1}{m} \right\}, \\ \overline{CR}_3 &= \left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2 W \\ \underline{CR}_3 &= V \cdot \min \left\{ \left(1 - \frac{1}{m^2}\right)W, \frac{1}{2}\left(1 - \frac{1}{m^2}\right)W + \frac{1}{2} \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right)\alpha, 1 - \frac{1-p}{pm}\ell_1 \right\} \right\}. \end{aligned}$$

To provide insights into the lower bound in Theorem 1, before we formally present the asymptotic CR of Algorithm 1, we provide an example to see how the CR of Algorithm 1 vary with  $m$  and  $p$ .

#### 4.2. A Simple Example: Evaluating the Lower Bound in Theorem 1

EXAMPLE 1. Consider a setting in which  $r_1 = 0.9$ , and  $r_2$  is uniformly drawn from the interval  $(0.5, 0.9)$ ; that is,  $r_2 \sim \text{Uniform}(0.5, 0.9)$ . Here, we would like to study how the true CR (which will be defined below) of Algorithm 1 and its lower bound in Theorem 1 change as a function of  $m$  and  $p$ . To do so, we consider two scenarios. In first scenario, we fix the sampling probability  $p$  at 0.3, and we let the number of resources  $m \in \{10, 35, \dots, 360, 385\}$ . For each value of  $m$ , we generate 300 instances where an instance is determined by a realized value of  $r_2 \sim \text{Uniform}(0.5, 0.9)$ . In the second scenario, we fix the number of resources  $m$  to be 200 and we let the sampling probability  $p \in \{0.1, 0.15, \dots, 0.55, 0.6\}$ . For each value of  $p$ , we generate 300 instances where again an instance is determined by the realized value of  $r_2$ .

For each instance of both scenarios, we then compute three quantities : (i) the true CR of Algorithm 1, (ii) the lower bound on CR of Algorithm 1, which is presented in Theorem 1, and (iii)  $1/(2 - r_2/r_1) = 1/(2 - \alpha)$ . The true CR of the algorithm is obtained by considering

all the possible values for  $h$  and  $\ell$ , and only focusing on ordered arrival sequences in which type 2 agents arrive first. Note that focusing on ordered arrival sequences is without loss of generality as shown in Lemma 1. The last quantity  $1/(2 - \alpha)$  is the optimal CR when  $r_i, i \in [2]$ , is known, but no sample information is available. We use this quantity as a benchmark.

Figure 1 shows the expected value of these three quantities versus  $m$  in the first scenario where we fix the sampling probability  $p = 0.3$ . Here, the expectation is with respect to  $r_2$ . The figure shows both the true CR and its lower bound in Theorem 1 improves as  $m$  increases. Our lower bound, however, gets tighter as  $m$  increases. Nonetheless, our lower bound is quite loose when  $m$  is small. This is mainly because of our loose lower bound for  $\text{CR}_1$ , where we recall that  $\text{CR}_1$  is the lower bound on the CR of the algorithm when  $h \leq \sqrt{m}$ . To characterize  $\text{CR}_1$ , we present a universal lower bound for any realization  $\mathbf{n}$ , which leads to a loose bound; see Lemma 2 and its proof.

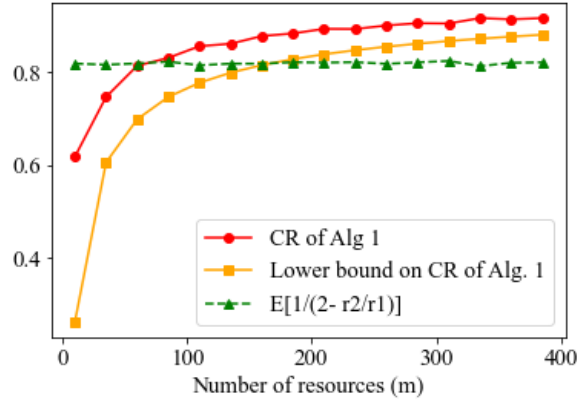
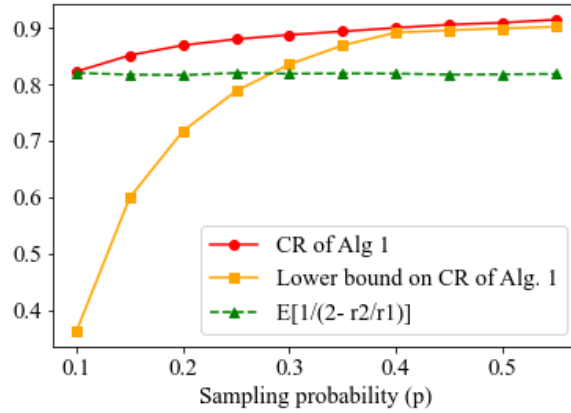
What is quite interesting is that the true CR of the algorithm exceeds the benchmark of  $1/(2 - \alpha)$  when  $m \geq 35$ . Recall that the benchmark is the optimal CR when  $r_i$ 's are known, but sample information is not available. This highlights that for large enough  $m$ , the value of the sample information outweighs the drawbacks of not knowing the true rewards.

Figure 2 shows the average of these three quantities versus  $p$  in the second scenario where  $m = 200$ . The figure shows that our lower bound gets tighter as  $p$  increases. When  $p \geq 0.4$ , we can find that the lower bound is almost equal to the true CR of Algorithm 1. In addition, with  $m = 200$ , even if the sampling probability is as small as 0.1, the true CR of Algorithm 1 is still higher than the benchmark. In addition, we observe that both the lower bound on the CR (given by Theorem 1) and the true CR of Algorithm 1 are larger in the high sampling probability case.

### 4.3. Asymptotic Competitive Ratio of Algorithm 1

Here, we present an asymptotic CR of Algorithm 1 as a function of  $m$  and  $p$  when  $m$  goes to infinity.

**PROPOSITION 2 (Asymptotic Competitive Ratio of Algorithm 1).** *Consider the model presented in Section 2, where the expected rewards of  $r_i, i \in \{1, 2\}$  is unknown to the decision-maker. As  $m$  goes to infinity, the worst-case CR of Algorithm 1 scales with  $1 - \Theta(1/(p\sqrt{m}))$ .*

Figure 1  $p = 0.3$ Figure 2  $m = 200$ 

**Figure 3** The CR of Algorithm 1, the lower bound on the CR of Algorithm 1 presented in Theorem 1, and the bound in Ball and Queyranne (2009) (i.e.,  $\mathbb{E}[1/(2 - \alpha)]$ ). Here,  $r_1 = 0.9$ , and  $r_2$  is uniformly drawn from the interval  $(0.5, 0.9)$ .

Proposition 2 is shown in Section C in the appendix. Proposition 2 shows that the CR of Algorithm 1 goes to one as  $m$  goes to infinity if  $p = \omega(1/\sqrt{m})$ . The asymptotic CR of the algorithm (i.e.,  $1 - \Theta(1/(p\sqrt{m}))$ ) is indeed tight for any  $p = \omega(1/\sqrt{m})$ ; see Section 5 for an upper bound on the CR of any deterministic and randomized algorithms when  $p = \omega(1/\sqrt{m})$ .<sup>11</sup> Note that even when  $p$  is a constant, it is not obvious why achieving asymptotic optimality is possible using Algorithm 1. Suppose that  $m$  is large and  $p$  is constant. If the adversary chooses a large market size (i.e., when both  $h$  and  $\ell$  are large), for a constant value of  $p$ , the number of agents in the sample set is large, providing significant information. However, the market size is chosen adversarially and hence it can be small either for type 1 or 2 agents, and in such a case, the sample information is not very

<sup>11</sup> For the case where  $p = O(1/\sqrt{m})$ , in Section 5, we further present a non-asymptotic upper bound. This bound also shows that when  $p = O(1/m)$ , no randomized and deterministic algorithms can be asymptotically optimal.



informative. (Recall that the asymptotic CR of Algorithm 1 is governed by region  $\mathcal{R}_1$ , where the number of type 1 agent is low.) Nevertheless, it turns out Algorithm 1 can very well handle the difficult cases where the market size is small, achieving an asymptotically optimal CR.

## 5. Upper Bound on the Competitive Ratio of All Deterministic and Randomized Algorithms

In our problem setting, the average rewards of agents are unknown, but the decision-maker can learn about them with the help of the sample information. In this section, we relax the problem a little bit by considering setting in which the decision-maker knows  $r_1$  and  $r_2$  in advance. We present an upper bound on the CR of any deterministic and randomized algorithm in the aforementioned relaxed setting when the number of resources  $m$  goes to infinity. Clearly, any upper bound for this relaxed setting is also a valid upper bound for our original setting in Section 2.

First, in Section 5.1, we consider the case where  $p = \omega(1/\sqrt{m})$ . For the case of  $p = \omega(1/\sqrt{m})$ , we present an upper bound on the CR of any deterministic and randomized algorithms under the relaxed setting. We show that the CR of any randomized and deterministic algorithm is upper bounded by  $1 - \Theta(1/(p\sqrt{m}))$  when  $p = \omega(1/\sqrt{m})$ , implying the asymptotic optimality of Algorithm 1; see Proposition 2.

Put differently, the asymptotic CR of our algorithm—which does not use any feedback throughout the allocation period to tune its estimates for  $r_i$ 's—matches the upper bound on the CR of any deterministic and randomized algorithm in the relaxed setting where  $r_i$ 's are known in advance. This implies that asymptotically, feedback during the allocation period does not add any value in improving the CR of algorithms. The fact that feedback throughout the allocation period is not useful is due to the adversarial nature of arrival sequences. Under adversarial arrivals, the decision-maker cannot control for what type of agents she receives feedback; this is governed by the adversarially-chosen order.

Second, in Section 5.2, we consider the case where  $p = O(1/\sqrt{m})$ . For this case where the sampling probability is quite small even for large values of  $m$ , we present a non-asymptotic upper bound. Our bound shows that when  $p = O(1/m)$ , the CR of any randomized and deterministic algorithm is upper bounded by  $1/(2 - \alpha)$  as  $m$  increases. This is mainly due to the lack of adequate sample information when  $p = O(1/m)$ .

### 5.1. Upper Bound for $p = \omega(1/\sqrt{m})$

Consider the case of  $p = \omega(1/\sqrt{m})$ . Under this case, in Theorem 2, we first present an upper bound of  $1 - \Theta(1/(p\sqrt{m}))$  on the CR of any deterministic algorithm under the relaxed setting. Second, in Theorem 3, we show that the same upper bound holds for any randomized algorithm.

**THEOREM 2 (Upper Bound of Any Deterministic Algorithm for  $p = \omega(1/\sqrt{m})$ ).** *Consider a special case of our model, presented in Section 2, where the expected rewards of  $r_i$ ,  $i \in \{1, 2\}$  is known to the decision-maker. When  $p = \omega(1/\sqrt{m})$ , as the number of resources  $m$  goes to infinity, any deterministic algorithm has the CR of at most  $1 - \Theta(1/(p\sqrt{m}))$ .*

The proof of Theorem 2 is presented in Section D. To show Theorem 2, we construct the following input family  $\mathbf{F}$ : Let  $\underline{h} = 0$  and  $\bar{h} = pm$ . The input family  $\mathbf{F}$  contains all  $(h, \ell)$  such that  $h \in [\underline{h}, \bar{h}]$  and  $\ell = \frac{10000 \cdot m}{p}$ . For any  $h \in [\underline{h}, \bar{h}]$ , we then denote  $I_h$  as a random arrival sequence under which  $n_2$  type 2 agents arrive followed by  $n_1$  type 1 agents, where we recall that  $n_1 \sim \text{Bin}(h, 1 - p)$ ,  $n_2 \sim \text{Bin}(\ell, 1 - p)$ . We characterize an upper bound on the CR of any deterministic algorithm under the family  $\mathbf{F}$ .

In this family, because  $\ell \gg m$ , we know there will be more than  $m$  type 2 agents showing up. Therefore, the number of type 2 agents in the sample does not impact the acceptance/rejection decisions. Then, given that the online arrival sequences  $I_h$  are all ordered, any deterministic algorithm has to decide about how many type 2 agents they accept provided that they observe  $s_1$  samples from type 1 agents. Put differently, any deterministic algorithm can be represented by a mapping that maps  $s_1$  to the number of type 2 agents it accepts. The proof of Theorem 2 then shows that the best CR under any such mapping is  $1 - \Theta(1/(p\sqrt{m}))$ . The main challenge in showing this result is characterizing the optimal mapping. To overcome this, instead of characterizing the optimal mapping, we first construct a specific mapping under which upon observing  $s_1$  type 1 agents, we accept  $m - s_1 \frac{1-p}{p}$  type 2 agents. Observe that  $s_1 \frac{1-p}{p}$  is the expectation of the number of type 1 agents who will arrive given that there are  $s_1$  type 1 agents in the sample. Lemma 13 shows that under this mapping, the CR is at most  $1 - \Theta(1/(p\sqrt{m}))$ . To complete the proof of Theorem 2, we then compare any other mappings with this specific mapping; see Lemma 14.

Next, we derive an upper bound on the CR of all randomized algorithms.

**THEOREM 3 (Upper Bound of Any Randomized Algorithms for  $p = \omega(1/\sqrt{m})$ ).** Consider a special case of our model, presented in Section 2, where the expected rewards of  $r_i$ ,  $i \in \{1, 2\}$  is known to the decision-maker. Then, when  $p = \omega(1/\sqrt{m})$ , as the number of resources  $m$  goes to infinity, any randomized algorithm has the CR of at most  $1 - \Theta(1/(p\sqrt{m}))$ .

To show Theorem 3, unfortunately, we cannot use Von Neuman/Yao principle Seiden (2000). This is because in our setting, even when the input  $(h, \ell)$  is realized, due to our sampling procedure, the online arrival sequence is still random. This is different from Von Neuman/Yao principle, because Von Neuman/Yao principle can only be applied to the model without any randomness. Nonetheless, in Lemma 15, we derive a result similar to the Von Neuman/Yao principle that can be applied to our setting. We then apply Lemma 15 by constructing a distribution over the input family  $\mathbf{F}$  introduced above. This leads to the desired upper bound. See Section E for the proof of Theorem 3.

## 5.2. Non-asymptotic Upper Bound for any $p$

In the previous section, we give a tight asymptotic upper bound for  $p = \omega(1/\sqrt{m})$ . There are two remaining questions: (i) Does there exist any non-asymptotic upper bound (even if it is loose)? (ii) What is the asymptotic upper bound when  $p$  goes to zero very fast as  $m$  grows? In this section, we try to answer these two questions.

The following theorem (Theorem 4) presents a non-asymptotic upper bound on the CR of any deterministic or randomized algorithms as a function  $p$  and  $m$ . The theorem further shows that when functions  $p = O(1/m)$ , achieving asymptotic optimality is not possible. With  $p = O(1/m)$ , the asymptotic upper bound is equal to  $\frac{1}{2-\alpha}$ , which is the bound in Ball and Queyranne (2009). This shows that when  $p$  goes to 0 very fast, the sample information is not adequate enough to make an impact. We note that the upper bound in Theorem 4 can be loose because to compute the upper bound, we consider a relaxed setting where the rewards are known to the decision-maker. Even so, in light of the impossibility results in Theorem 4, in Section J, we present a simple algorithm whose CR is 0.5 for any value of  $p$  and  $m$ . While this CR does not match the upper bound in Theorem 4, the CR is tight when  $r_2 \rightarrow 0$  and  $p = O(1/m)$ .

**THEOREM 4 (Non-Asymptotic Upper Bound of Any Algorithms for any  $p$ ).** Any deterministic or randomized algorithm cannot achieve a CR better than

$\frac{m}{m + \min\{\tilde{h}, m\}(1-\alpha)} + \frac{1}{\tilde{h}^2}$ , where  $\tilde{h}$  is the smallest integer such that  $\tilde{h}p + \sqrt{\tilde{h} \log(\tilde{h})} \geq 1$ , and  $\alpha = r_2/r_1$ . In addition, when  $p = O(1/m)$ , no algorithm can obtain a CR better than  $\frac{1}{2-\alpha}$  as  $m \rightarrow \infty$ .

The proof is in Section F. Figure 4 depicts the upper bound in Theorem 4 versus  $m$  with  $p = 1/\sqrt{m}$  and  $p = 1/m$ , respectively. Here,  $r_1 = 0.9$  and  $r_2 = 0.5$ . As a benchmark, in both figures, we also include  $1/(2 - r_2/r_1)$ , which is the bound in Ball and Queyranne (2009). As expected, we observe that when  $p = 1/m$ , the upper bound converges to the benchmark. When  $p = 1/\sqrt{m}$ , however, the upper bound goes to 1 because this bound is loose.

To show Theorem 4, we construct an input family  $\mathbf{F}_1$  as follows: The input family  $\mathbf{F}_1$  contains all  $(h, \ell)$  such that  $h \in [0, \lfloor m/p \rfloor]$  and  $\ell = \lceil \frac{10000 \cdot m}{p} \rceil$ . For any  $h \in [0, \lfloor m/p \rfloor]$ , we then denote  $I_h$  as an ordered random arrival sequence under which  $n_2 \sim \text{Bin}(\ell, 1 - p)$  type 2 agents arrive followed by  $n_1 \sim \text{Bin}(h, 1 - p)$  type 1 agents. We then show that no algorithm can do well simultaneously on arrival sequences  $I_h$  for any  $h \in [0, \tilde{h}]$ , where  $\tilde{h}$  is the smallest integer such that  $\tilde{h}p + \sqrt{\tilde{h} \log(\tilde{h})} \geq 1$ . (Asymptotically,  $\tilde{h} = \Theta(1/p)$ ) Under these arrival sequences, with high probability,  $s_1 = 0$ . But these sequences vary a lot in terms of the number of type 1 agents. Yet it is not possible to distinguish them using the sample information.

## 6. A Case Study on COVID-19 Dataset

In this section, we do a case study in which we apply Algorithm 1 to the COVID-19-associated hospitalizations problem in the US. We use Algorithm 1 to allocate hospital resources (e.g., a medicine) to arriving COVID-19 patients sequentially. This case study allows us to evaluate Algorithm 1 in real-world inspired settings. It further allows us to see how robust the algorithm is to the estimation errors in the sampling probability  $p$ . We observe that the average CR of the algorithm even for small values of  $m$  ( $m < 50$ ) is at least 0.88, and the algorithm maintains its performance when it does not know the true value of the sampling probability. Further, we demonstrate the value of the sample information and running a test period by comparing our algorithm with a benchmark that does not use the sample information.

**Dataset.** Here, we use the ‘‘Laboratory-Confirmed COVID-19-Associated Hospitalizations’’ dataset. This dataset contains the number of bi-weekly cases of COVID-19-associated hospitalizations in the US from March 7<sup>th</sup>, 2020 to February 5<sup>th</sup>, 2022 across

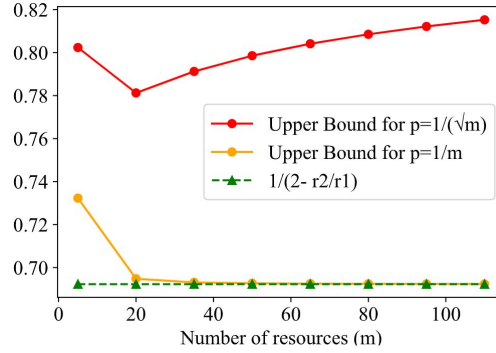


Figure 4

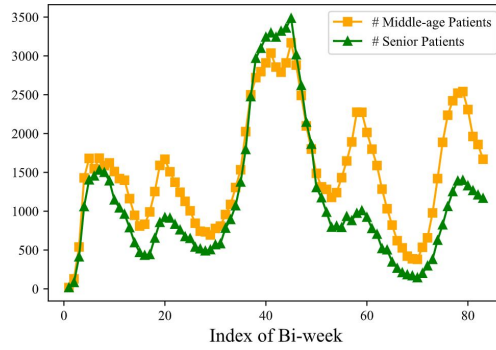


Figure 5

**Figure 6** (a) Upper bound in Theorem 4 for  $p = 1/\sqrt{m}$  and  $p = 1/m$ , and the bound in Ball and Queyranne (2009). Here,  $r_1 = 0.9$ , and  $r_2 = 0.5$ . (b) Number of middle-age patients and senior patients over weeks

five age groups of patients: 0-4 years, 5-17 years, 18-49 years, 50-64 years, and 65+ years. Because the number of 0-17 years old patients is small, we discard them in our analysis. We then consider two groups/types of patients: middle-age patients (18-64 years), and senior patients (65+ years). In our studies, the middle-age patients are considered to be the low-reward agents (i.e., type 2 is our setting), and senior patients are considered to be high-reward agents (i.e., type 1 in our setting). This assumption is partly motivated by high death-rate of senior patients when contracting COVID-19 (The-New-York-Times 2022).

**Simulation setup.** In our studies, given the description of our dataset, we consider the allocation problem of a resource (e.g., certain medicine) over the course of two weeks, where each resource can be assigned to at most one patient and each patient needs one unit of the resource. For each such period, we determine the number of hospitalized patients of the two aforementioned types (i.e.,  $h$  and  $\ell$ ) using our dataset. (Note that we have 83 periods in our dataset.) The value of  $h$  and  $\ell$  over 83 periods can be found in Figure 5. At the beginning of each of the periods, we observe the (realized) effectiveness of the resource for a sample of

hospitalized patients, where each of the  $(h + \ell)$  hospitalized patients falls into the sample with probability  $p$ . These sampled patients are, for example, the patients that arrive at the beginning of the two-week period during which rationing the resource has not started yet. Let  $n_1$  and  $n_2$  be the number of hospitalized patients of types 1 and 2 who did not fall into the sample. We assume that the order over these  $n_1 + n_2$  patients are completely random. That is, we consider uniform permutations over these patients, modeling more realistic scenarios where the order over patients are not chosen adversarially.

We consider 9 problem classes where each problem class is determined by two parameters  $(p, \gamma)$ . Here,  $p \in \{0.1, 0.15, 0.2\}$  is the sampling probability, and  $\gamma \in \{0.3, 0.5, 0.7\}$  determines the scarcity of the resource. In particular, we set  $m = \gamma \cdot (\hat{h} + \hat{\ell})$ , where  $\hat{h}$  and  $\hat{\ell}$  are respectively the number of type 1 and type 2 patients in the previous period (i.e., the last two weeks). By setting  $m$  in this way, we would like to capture the hospital's inventory planning decisions based on the most recent observed demand. Further, in choosing the values for the number of resources  $m$ , we let  $\gamma$  be less than one. This allows us to model scenarios where resources are scarce. When resources are not scarce, the problem is not challenging.

For each problem class, we generate 1000 instances, where each instance is determined by  $(n_1, n_2)$  and an order over  $n_1 + n_2$  patients. For each of the 9 problem classes, we evaluate the performance of four algorithms: (i) Algorithm 1 that has access to the true value of  $p$ , (ii) an algorithm that does not use sample information (which we will define later in this section), (iii) an adaptive version of Algorithm 1 that has access to the true value of  $p$  (see Section 4 for the description of this algorithm), and (iv) Algorithm 1 that has only access to the noisy estimate of  $p$ , denoted by  $\hat{p}$ , and uses  $\hat{p}$  in place of  $p$ . Here,  $\hat{p} \sim \text{Uniform}(\underline{p}, \bar{p})$ ,  $\mathbb{E}[\hat{p}] = p$ , and standard error of  $(\hat{p})$  is set to  $0.3p$ . For example, if  $p = 0.1$ , we have  $\underline{p} \approx 0.05$  and  $\bar{p} \approx 0.15$ .

The second algorithm is similar to Algorithm 1 in the sense that it also assigns protection levels to either type 1 or type 2 agents. But, unlike Algorithm 1, it does not use the sample information to set the protection level and decide what type of agents to protect. This algorithm, which is parameterized by  $x$ , works as follows: with a probability  $1/2$ , it assigns a protection level of  $x \in \{0, 1, \dots, m\}$  to type 1 agents. Otherwise, it assigns a protection level of  $x$  to type 2 agents. The optimal (CR-maximizing) value for  $x$  can be calculated as

a function of  $\alpha = r_2/r_1$  and  $m$ .<sup>12</sup> Although without the sample information, rewards  $r_1$  and  $r_2$ , and hence  $\alpha$  are unknown, to give the algorithm an extra benefit, in our case study, we set  $x$  optimally. We highlight that we use this algorithm in our case study to shed light on the value of the sample information.

Finally, we note that in all of our evaluations, we assume that  $r_1 = 0.6$  and  $r_2 = 0.2$ . We also try to let  $r_i$ 's take other values: the take-away messages of our case study do not change.

**Performance evaluation.** Table 1 presents the worst-case and average CR of the four algorithms across the 82<sup>13</sup> allocation periods for each of our 9 problem classes, where in each allocation period, the CR is the worse case (minimum) reward over optimal clairvoyant reward among the 1000 instances. Then, the worst-case and average CR reported in Table 1 are computed by taking the minimum and average over the CR of 82 allocation periods.

Interestingly, for all problem classes, the average CR of our algorithm and its noisy version is between 0.88 and 0.97 while the average CR of the algorithm that does not use the sample information can be as low as 0.66. We also observe that Algorithm 1 and the adaptive version have almost the same CR. This is because, under the adaptive algorithm, in average (in worst case respectively), we change our decision about whom to protect less than 1 time (twice respectively).

In addition, for all four algorithms, if we fix  $\gamma$ , we observe that by increasing the sampling probability  $p$ , both the average and worst-case CR increase. This observation is consistent with the result in Proposition 2. However, if we fix the sampling probability  $p$ , there is no clear pattern of how the CR changes when  $\gamma$  increases.

Overall, we observe that Algorithm 1 with true value of  $p$  and Algorithm 1 with noisy  $\hat{p}$  significantly outperform the naive algorithm that does not take advantage of the sample information.

This suggests the importance using the sample information even when we do not have access to the true value of  $p$ .

One surprising result is that: among all 9 problem classes and 4 algorithms, the worst CR happens in the second allocation period among the 82 periods. (Recall that we did

<sup>12</sup> The optimal value for  $x$  is  $\frac{2-2\alpha}{2-\alpha}m$ .

<sup>13</sup> Since in each two-week period, we set  $m$  as a function of demand in the previous two-week period, we cannot have the results for the first two-week period.

not analyze the first allocation period because how we set  $m$ .) From Figure 5, we can find that the second allocation period has the smallest number of type 1 and 2 patients. Hence, this observation is consistent with the insights from our theoretical result, which show that when the market size is small, the CR is also small.

**Table 1** The CR of four algorithms in Section 6. The standard error of average CRs of the four algorithms is less than 0.001.

	Value of $p$	p=0.1			p=0.15			p=0.2		
	Value of $\gamma$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 0.7$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 0.7$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 0.7$
Alg. 1 with true $p$	Avg. CR	0.977	0.953	0.961	0.978	0.962	0.968	0.978	0.965	0.973
	Worst CR	0.791	0.763	0.811	0.808	0.787	0.824	0.811	0.771	0.862
Alg that doesn't use samples	Avg. CR	0.656	0.693	0.838	0.657	0.711	0.872	0.671	0.737	0.912
	Worst CR	0.567	0.573	0.767	0.589	0.649	0.810	0.593	0.671	0.860
Adaptive version of Alg. 1 with true $p$	Avg. CR	0.979	0.955	0.962	0.980	0.962	0.968	0.979	0.967	0.973
	Worst CR	0.800	0.766	0.811	0.837	0.792	0.831	0.866	0.791	0.864
Alg. 1 with noisy $p$	Avg. CR	0.953	0.892	0.888	0.944	0.889	0.910	0.931	0.881	0.917
	Worst CR	0.790	0.723	0.775	0.817	0.800	0.811	0.799	0.781	0.849

## 7. Extensions and Discussions

In this section, we discuss some extensions of our model. In Section 7.1, we study a setting where the number of resources during the test period is limited, and in Section 7.2, we investigate a strong clairvoyant benchmark that knows the realized reward of agents. In Section 7.3, we discuss how to modify Algorithm 1 when the sampling probability is different across different types of agents. We further investigate how the modified version of Algorithm 1 performs in the worst case. In Section 7.4, we present an extension of Algorithm 1 when there are more than two types of agents.

### 7.1. Capacity Constraints during the Test Period

So far, we have assumed that there is no capacity constraint during the test period in the sense that any agent in the test period gets one unit of the resource. However, in practice, decision-makers may limit the number of available resources during the test period due to the cost of testing the market. Here, we show that when the number of available resources during the test period is  $\omega(\sqrt{m})$  (e.g.,  $\Theta(m^{1/2+\epsilon})$  for any  $\epsilon > 0$ ) the same asymptomatic CR, presented in Proposition 2, continues to hold.

Let  $m_t$  and  $m_a = m$  be the number of available resources during the test and allocation periods, respectively. Further, recall that  $s_i$ ,  $i \in \{1, 2\}$ , is the number of agents of type  $i$  in the sample, where we assume that  $s_i$ 's are known to the decision-maker when enforcing the



capacity constraint.<sup>14</sup> To enforce the capacity constraint, when  $s_1 + s_2 \leq m_t$ , clearly, we can accept all the (sampled) agents in the test period. On the other hand, when  $s_1 + s_2 > m_t$  with  $s_1, s_2 > m_t/2$ , we accept  $m_t/2$  agents of type  $i \in \{1, 2\}$ . Finally, when  $s_1 + s_2 > m_t$  with either  $s_1$  or  $s_2$  (not both) is greater than  $m_t/2$ , we accept a total of  $m_t$  agents, while making sure all the agent of the type with a lower number of samples is accepted. At the end of the test period, for simplicity, we assume that the leftover units do not carry over to the allocation period.

The following proposition presents the asymptotic CR of Algorithm 1 with a slight modification: in calculating the empirical average awards  $\widehat{r}_i$ 's, instead of using  $s_i$ 's, we use the number of accepted agents of type  $i$  at the end of the test period.

**PROPOSITION 3 (Capacity Constraint during the Test Period).** *Consider the model presented in Section 7.1 where the number of resources during the test period is at most  $m_t$ . Then, when  $m_t = \omega(\sqrt{m})$ , the worst-case CR of a modified version of Algorithm 1, stated above, scales with  $1 - \Theta(1/(p\sqrt{m}))$  as  $m$  goes to infinity. Here,  $m$  is the number of resources during the allocation period.*

The proof can be found in Section G. The proof is similar to that of Theorem 2. As the main difference, we need to re-evaluate the probability of good event which is needed in the proof of region  $\mathcal{R}_3$ . Proposition 3 shows that even when the number of resources during the test period is quite small, Algorithm 1 maintains its good performance.

## 7.2. A Different Notion of Competitive Ratio using Realized Rewards

In our setting, we compared the performance of our algorithm w.r.t. an optimal clairvoyant benchmark that knows the arrival sequence and the expected reward of the agents in advance, but does not know the realized rewards of the agents. See Equation (1) for the definition of our CR notion. This benchmark is indeed consistent with the benchmark used in Ball and Queyranne (2009). Nevertheless, one may wonder how our algorithm performs against a stronger benchmark that also knows the realized reward of the agents, denoted by  $\mathcal{R} = \{R_t\}_{t \in [|I|]}$ . (Here, when  $z_t = i$ , the realized reward  $R_t \sim \text{Ber}(r_i)$ , and the realized reward is not known at the time of the decision.) Such a clairvoyant benchmark only accepts agents with realized reward of one, where we recall that the realized reward of agents is either zero or one.

<sup>14</sup> That is, all agents who respond to the outreach program arrive first. After their arrivals, we decide about allocating resources to them.

Let  $\text{OPT-RE}(\mathcal{R}, I)$  be the maximum realized reward under  $\mathcal{R}$  and the arrival sequence  $I$ , where “RE” stands for “realized.” Further, let  $\text{REW-RE}_{\mathcal{A}}(\mathcal{R}, I, \psi)$  be the realized reward of an algorithm  $\mathcal{A}$  under  $\mathcal{R}$ , arrival sequence  $I$ , and the sample information  $\psi$ . Then, the CR of an algorithm  $\mathcal{A}$  w.r.t. the strong benchmark  $\text{OPT-RE}(\mathcal{R}, I)$ , denoted by  $\text{CR-RE}$ , is defined as

$$\text{CR-RE}_{\mathcal{A}} = \inf_I \mathbb{E} \left[ \inf_{\mathcal{R}} \frac{\mathbb{E}[\text{REW-RE}_{\mathcal{A}}(\mathcal{R}, I, \psi)]}{\text{OPT-RE}(\mathcal{R}, I)} \right], \quad (8)$$

where the inner expectation is w.r.t. any randomness in algorithm  $\mathcal{A}$ , and the outer expectation is w.r.t. the sample information and realized rewards  $\mathcal{R}$ . We refer to  $\text{CR-RE}_{\mathcal{A}}$  as the *realized CR* of algorithm  $\mathcal{A}$ .

In this section, we first show that for any randomized and deterministic algorithm  $\mathcal{A}$ , the realized CR cannot exceed  $r_2$  even when expected rewards are known. See Proposition 4. This highlights that it is very challenging to beat the realized benchmark  $\text{OPT-RE}(\mathcal{R}, I)$ . We then revisit Example 1 in which we evaluate Algorithm 1 using the realized CR. The example shows that the realized CR of Algorithm 1 converges to  $r_2$  as  $m$  increases, implying Algorithm 1 might be also optimal in terms of the realized CR.

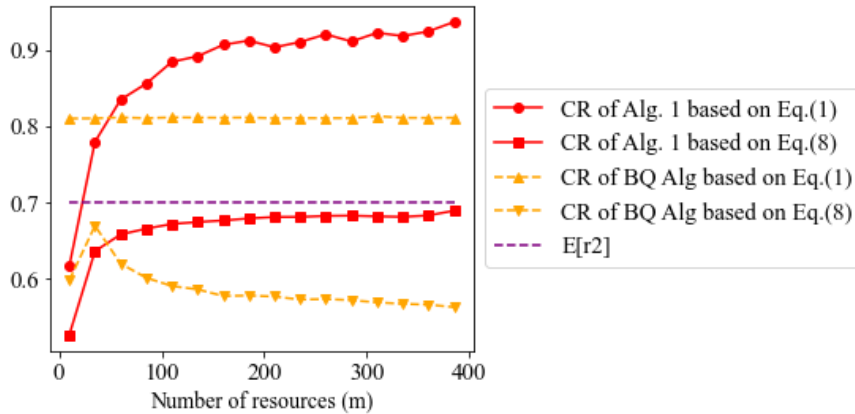
**PROPOSITION 4 (Upper Bound on the Realized CR).** *Consider a relaxed version of our setting where the expected rewards are known, but we still have access to sample information  $\psi$  that reveals some information about the market size  $h$  and  $\ell$ . Then, the realized CR of any deterministic or randomized algorithm (per Equation (8)) is at most  $r_2$ .*

The proof can be found in Appendix H. By Proposition 4,  $r_2$  is an upper bound among all deterministic and randomized algorithms for our setting where the expected rewards are unknown. In light of the lower bound in Proposition 4, next we revisit Example 1 where we evaluate Algorithm 1 and the algorithm of Ball and Queyranne (2009) (called the BQ algorithm) using the notion of realized CR.

**EXAMPLE 2 (REVISITING EXAMPLE 1).** Consider the same setting in Example 1 with the sampling probability  $p = 0.3$ ,  $m \in \{10, 35, \dots, 360, 385\}$ ,  $r_1 = 0.9$ , and  $r_2 \sim \text{Uniform}(0.5, 0.9)$ . For this setting, we compute four quantities: (i) the CR of Algorithm 1 per Equation 1, (ii) the realized CR of Algorithm 1 per Equation (8), (iii) the CR of the BQ algorithm (i.e.,  $1/(2 - r_2/r_1) = 1/(2 - \alpha)$ ), (iv) the realized CR of the BQ algorithm.

(Recall that (i) and (iii) are computed in Example 1 and here they are just used for comparison.) Note that the BQ algorithm is evaluated in a setting where the expected rewards are known. To compute these quantities (i.e., quantities (i), (ii), and (iv)), we considered all the possible values for  $h$  and  $\ell$ , any arrival order, and the realization of reward.

Figure 7 shows the expected value of these four quantities versus  $m$  with  $p = 0.3$ . Here, the expectation is with respect to  $r_2$ . The figure shows as expected, the realized CR of Algorithm 1 and the BQ algorithm is much smaller than their CR per Equation (1). As another important observation, the realized CR of Algorithm 1 goes to  $\mathbb{E}[r_2]$ , which is the upper bound shown in Proposition 4. In addition, in terms of realized CR, Algorithm 1 outperforms the BQ algorithm for  $m \geq 60$ , which is consistent with what we observed in Example 1.



**Figure 7** The CR and the realized CR of Algorithm 1 and the BQ algorithm, where the CR is computed using Equation (1) and the realized CR is computed using Equation (8). Here,  $p = 0.3$ ,  $r_1 = 0.9$ , and  $r_2$  is uniformly drawn from the interval  $(0.5, 0.9)$ .

### 7.3. Heterogeneous Sampling Probability

In our setting, we assumed that any of the  $h + \ell$  agents reacts to the outreach program with probability  $p$ . In practice, however, different types of agents may react to the outreach program differently. To capture that, here we study a setting where type  $i \in \{1, 2\}$  agents react to the outreach program (i.e., falls into the sample set) with probability  $p_i$ .

For this setting, let us consider a slightly modified version of Algorithm 1. In this modified algorithm, once  $\hat{r}_1 > \hat{r}_2$  ( $\hat{r}_1 \leq \hat{r}_2$  respectively), the algorithm assigns a protection level of  $\min\{m, s_1 \frac{1-p_1}{p_1}\}$  ( $\min\{m, s_2 \frac{1-p_2}{p_2}\}$  respectively) to type 1 (type 2 respectively). The following theorem presents a lower bound on the CR of this modified algorithm.

**PROPOSITION 5 (Heterogeneous  $p$ ).** *Consider the model presented in Section 2, where the expected rewards of  $r_i$ ,  $i \in \{1, 2\}$  is unknown to the decision-maker. Let the sampling probability of type 1, 2 agents be  $p_1, p_2$  respectively. Then, the asymptotic CR of the modified algorithm, defined above, equals to*

$$CR = 1 - \Theta \left( \max \left\{ \frac{1}{p_1 \sqrt{m}}, \frac{1}{\min\{p_1, p_2\}^2 m} \right\} \right).$$

The proof can be found in Section I. From the expression of the CR, we can find that if  $p_1 < p_2$ , then  $\Theta(\frac{1}{p_1 \sqrt{m}})$  is the dominant term if and only if  $p_1 = \omega(1/\sqrt{m})$ . That is, in this case, the asymptotic CR is  $1 - \Theta(\frac{1}{p_1 \sqrt{m}})$ . If  $p_1 \geq p_2$ , then  $\Theta(\frac{1}{p_1 \sqrt{m}})$  is the dominant term if and only if  $\frac{p_1}{p_2} = o(\sqrt{m})$ .

#### 7.4. Beyond Two Types of Agents

So far, we have assumed that there are only two types of agents, and under this assumption, we designed an asymptotically optimal algorithm (Algorithm 1). Here, we show how to generalize Algorithm 1 to a setting with  $k \geq 3$  type. In the generalized algorithm, similar to Algorithm 1, we first estimate the average reward of each type of agents using the sample information. We then decide about an order over the agents using the estimated rewards, and follow a nesting protection level policy given the order; see Algorithm 2. Consider an order  $(j_1, j_2, \dots, j_k)$  over types and protection levels  $x_1, x_2, \dots, x_k$ . Given this order and protection levels  $x_1, x_2, \dots, x_k$ , the nested protection policy works as follows. It accepts the arriving agent of type  $j_i$  if and only if (i) there is resource left (ii) the total number of accepted agents of type  $j_q$  for  $q = i, i+1, \dots, k$  is less than  $m - x_{i-1}$ . Here, we set  $x_0 = 0$ .

---

#### Algorithm 2 Online Resource Allocation Algorithm for $k$ types of Agents

---

**Input:** The number of resources  $m$  and sample information  $\psi = (\mathbf{s}, \boldsymbol{\rho})$ , where  $\mathbf{s} = (s_1, s_2, \dots, s_k)$ ,  $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_k)$ , and for any  $i \in [k]$ ,  $\boldsymbol{\rho}_i = \{\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,s_i}\}$ .

1. If the number of samples for type  $i \in [k]$ , i.e.,  $s_i$ , is positive, define  $\hat{r}_i = \frac{\sum_{j=1}^{s_i} \rho_{i,j}}{s_i}$  as an estimate of the expected reward of type  $i \in [k]$ . Otherwise,  $\hat{r}_i \sim \text{uniform}(0, 1)$ .
  2. Sort  $\hat{r}_i$  for  $i \in [k]$  from the largest to the smallest, i.e.,  $\hat{r}_{j_1} \geq \hat{r}_{j_2} \dots \geq \hat{r}_{j_k}$ .
  3. For  $i \in [k-1]$ , we give  $x_i = \min\{m, \frac{1-p}{p} \sum_{q=1}^i s_{j_q}\}$  protection level to type  $j_1, j_2, \dots, j_i$  agents and run a nested protection level policy with the order  $j_1, j_2, \dots, j_k$  and protection levels  $x_i, i \in [k]$ .
- 

We believe that Algorithm 2 is also asymptotically optimal although we find it hard to characterize a lower bound on its CR due to a large number of cases/regions that one needs to consider in its analysis. To demonstrate its good performance though, in the following we present an example in which we evaluate Algorithm 2 numerically.

EXAMPLE 3. Consider a setting in which  $r_1 = 0.9$ ,  $r_2 \sim \text{Uniform}(0.5, 0.9)$ , and  $r_3 \sim \text{Uniform}(0.1, 0.5)$ . Figure 8 shows the average of (i) the true CR of Algorithm 2, and (ii)  $1/(3 - r_3/r_2 - r_2/r_1)$  versus  $m$  when  $p = 0.3$ . Here, the average is taken w.r.t. the randomness in  $r_2$  and  $r_3$  in our 300 randomly generated instances. Note that  $1/(3 - r_3/r_2 - r_2/r_1)$  is the optimal CR in Ball and Queyranne (2009) for a setting with three types and known rewards. We use  $1/(3 - r_3/r_2 - r_2/r_1)$  as a benchmark to measure the value of the sample information. To compute the true CR, we take all  $h, \ell \geq 0$ , and for each  $(h, \ell)$ , we simulate 1000 instances with random order (the first instance is fixed to be the ordered instance). By taking the minimum CR among all sequences described above, we compute the true CR.

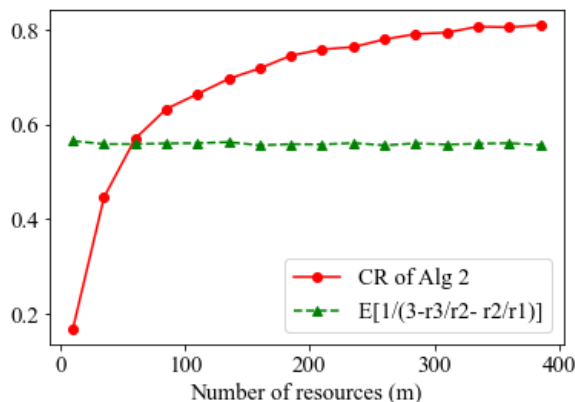
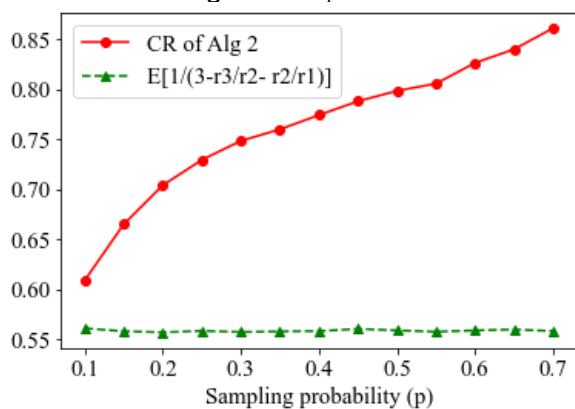
Figure 8 shows that by increasing  $m$ , the CR of Algorithm 2 improves. In addition, the true CR of the algorithm exceeds the benchmark when  $m \geq 60$ . This highlights that even in the 3-type case, for large enough  $m$ , the value of the sample information outweighs the drawbacks of not knowing the true rewards.

Figure 9 shows the average of the true CR of Algorithm 2 and the benchmark versus  $p$  when  $m = 200$ . Interestingly, we observe that when  $m = 200$ , even if the sampling probability is as low as 0.1, the true CR is much larger than the benchmark.

## 8. Concluding Remarks and Future Directions

In this paper, we consider an online resource allocation problem in which the decision-maker has uncertainty about the arrival process, as well as, the obtained rewards upon the allocation. The decision-maker has access to the sample information that is often acquired through an initial test period. We study how to optimally exploit the sample information that provides partial knowledge about the arrival process and rewards. We propose a protection-level algorithm that achieves the competitive ratio of  $1 - \Theta(1/(p\sqrt{m}))$  and show that the obtained competitive ratio is asymptotically tight in terms of both the initial number of resources  $m$  and the sampling probability  $p$ .

Our result shows that the sample information can be significantly beneficial, and hence opens up several new directions for future research. One natural direction is to study how to optimize the number of resources  $m$  upon receiving the sample information. While in many settings, due to a long lead production time, it is not possible to react to the sample information to optimize the inventory decisions, for some other settings when

Figure 8  $p = 0.3$ Figure 9  $m = 200$ 

**Figure 10** The CR of Algorithm 2, and the benchmark in Ball and Queyranne (2009). Here,  $r_1 = 0.9$ ,  $r_2$  is uniformly drawn from the interval  $(0.5, 0.9)$ , and  $r_3$  is uniformly drawn from the interval  $(0.1, 0.3)$ .

products/resources are produced domestically, the decision-makers can better plan their inventory decisions using the sample information. Another research question of interest is to study how to generalize our setting to the network revenue management problem where there are multiple types of agents and multiple types of resources. The network revenue management problem under adversarial arrival settings does not admit a constant competitive ratio. Thus, it is interesting to explore if one can even obtain a constant competitive ratio under adversarial arrival setting in the presence of sample information.

## References

- Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. 2014. A dynamic near-optimal algorithm for online linear programming. *Operations Research* 62, 4 (2014), 876–890.
- Kannapha Amaruchkul, William L Cooper, and Diwakar Gupta. 2007. Single-leg air-cargo revenue management. *Transportation science* 41, 4 (2007), 457–469.

- Antonios Antoniadis, Themis Gouleakis, Pieter Kleer, and Pavel Kolev. 2020. Secretary and online matching problems with machine learned advice. *arXiv preprint arXiv:2006.01026* (2020).
- Victor F Araman and René Caldentey. 2011. Revenue management with incomplete demand information. *Encyclopedia of Operations Research*. Wiley (2011).
- Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002a. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning* 47 (2002), 235–256.
- P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. 2002b. The non-stochastic multi-armed bandit problem. *SIAM journal of computing* 32 (2002), 48–77.
- Michael O Ball and Maurice Queyranne. 2009. Toward robust revenue management: Competitive analysis of online booking. *Operations Research* 57, 4 (2009), 950–963.
- Santiago Balseiro, Negin Golrezaei, Mohammad Mahdian, Vahab Mirrokni, and Jon Schneider. 2019. Contextual bandits with cross-learning. In *Advances in Neural Information Processing Systems*. 9676–9685.
- Santiago Balseiro, Christian Kroer, and Rachitesh Kumar. 2022. Single-Leg Revenue Management with Advice. *arXiv preprint arXiv:2202.10939* (2022).
- Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. 2020. The best of many worlds: Dual mirror descent for online allocation problems. *arXiv preprint arXiv:2011.10124* (2020).
- Daniel Berend and Aryeh Kontorovich. 2013. A sharp estimate of the binomial mean absolute deviation with applications. *Statistics & Probability Letters* 83, 4 (2013), 1254–1259.
- Omar Besbes and Assaf Zeevi. 2012. Blind network revenue management. *Operations research* 60, 6 (2012), 1537–1550.
- Jinzhi Bu, David Simchi-Levi, and Yunzong Xu. 2020. Online pricing with offline data: Phase transition and inverse square law. In *International Conference on Machine Learning*. PMLR, 1202–1210.
- Niv Buchbinder, Kamal Jain, and Joseph Seffi Naor. 2007. Online primal-dual algorithms for maximizing ad-auctions revenue. In *European Symposium on Algorithms*. Springer, 253–264.
- Ningyuan Chen and Shuoguang Yang. 2022. Bridging Adversarial and Nonstationary Multi-armed Bandit. *arXiv preprint arXiv:2201.01628* (2022).
- Qinyi Chen, Negin Golrezaei, and Djallel Bouneffouf. 2021b. Dynamic Bandits with Temporal Structure. *Available at SSRN 3887608* (2021).
- Xingxing Chen, Jacob Feldman, Panos Kouvelis, and Seung Hwan Jung. 2021a. Approximation schemes for the joint inventory selection and online resource allocation problem. *Production and Operations Management* (2021).
- José Correa, Andrés Cristi, Laurent Feuilloley, Tim Oosterwijk, and Alexandros Tsigonias-Dimitriadis. 2021. The secretary problem with independent sampling. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2047–2058.

- Nikhil R Devanur and Thomas P Hayes. 2009. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proceedings of the 10th ACM conference on Electronic commerce*. 71–78.
- Hossein Esfandiari, Nitish Korula, and Vahab Mirrokni. 2015. Online allocation with traffic spikes: Mixing adversarial and stochastic models. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. 169–186.
- Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S Mirrokni, and Cliff Stein. 2010. Online stochastic packing applied to display ad allocation. In *European Symposium on Algorithms*. Springer, 182–194.
- Kris Johnson Ferreira, David Simchi-Levi, and He Wang. 2018. Online network revenue management using thompson sampling. *Operations research* 66, 6 (2018), 1586–1602.
- Guillermo Gallego, Lin Li, and Richard Ratliff. 2009. Choice-based EMSR methods for single-leg revenue management with demand dependencies. *Journal of Revenue and Pricing Management* 8, 2 (2009), 207–240.
- Negin Golrezaei, Hamid Nazerzadeh, and Paat Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Science* 60, 6 (2014), 1532–1551.
- Negin Golrezaei and Evan Yao. 2021. Online Resource Allocation with Time-Flexible Customers. *arXiv preprint arXiv:2108.03517* (2021).
- Dawsen Hwang, Patrick Jaillet, and Vahideh Manshadi. 2021. Online resource allocation under partially predictable demand. *Operations Research* (2021).
- Stefanus Jasin. 2015. Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research* 63, 4 (2015), 909–915.
- Billy Jin and Will Ma. 2022. Online Bipartite Matching with Advice: Tight Robustness-Consistency Tradeoffs for the Two-Stage Model. *arXiv preprint arXiv:2206.11397* (2022).
- Haim Kaplan, David Naori, and Danny Raz. 2022. Online Weighted Matching with a Sample. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1247–1272.
- Yingjie Lan, Huina Gao, Michael O Ball, and Itir Karaesmen. 2008. Revenue management with limited demand information. *Management Science* 54, 9 (2008), 1594–1609.
- Silvio Lattanzi, Thomas Lavastida, Benjamin Moseley, and Sergei Vassilvitskii. 2020. Online scheduling via learned weights. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 1859–1877.
- K Littlewood. 1972. Forecasting and control of passenger bookings. *AGIFORS Sympos. Proc* 12 (1972), 95–117.
- Thodoris Lykouris and Sergei Vassilvitskii. 2018. Competitive caching with machine learned advice. In *International Conference on Machine Learning*. PMLR, 3296–3305.



- Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. 2007. Adwords and generalized online matching. *Journal of the ACM (JACM)* 54, 5 (2007), 22–es.
- SV Nagaev and VI Chebotarev. 2011. On the bound of proximity of the binomial distribution to the normal one. In *Doklady Mathematics*, Vol. 83. Springer, 19–21.
- Rad Niazadeh, Negin Golrezaei, Joshua R Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. 2021. Online learning via offline greedy algorithms: Applications in market design and optimization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*. 737–738.
- Georgia Perakis and Guillaume Roels. 2010. Robust controls for network revenue management. *Manufacturing & Service Operations Management* 12, 1 (2010), 56–76.
- Pfizer. 2022. Guide to Clinical Trials: Potential Breakthroughs Put to the Test. <https://www.pfizer.com/science/clinical-trials/guide> (2022).
- productreviewmom.com. 2022. Get Paid to Test and Review Products at Home Free in 2022. <https://www.productreviewmom.com/2013/06/how-can-i-get-paid-to-test-products.html> (2022).
- Steven S Seiden. 2000. A guessing game and randomized online algorithms. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. 592–601.
- The-New-York-Times. 2022. As U.S. Nears 800,000 Virus Deaths, 1 of Every 100 Older Americans Has Perished. [www.nytimes.com/2021/12/13/us/covid-deaths-elderly-americans.html](http://www.nytimes.com/2021/12/13/us/covid-deaths-elderly-americans.html) (2022).
- W. R. Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25 (1933), 285 – 294.
- Bart Van Parys and Negin Golrezaei. 2020. Optimal learning for structured bandits. *Available at SSRN 3651397* (2020).
- Andrew Chi-Chin Yao. 1977. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*. IEEE Computer Society, 222–227.

## Appendix A: Proof of Proposition 1

We show the results for the three scenarios respectively. Recall that in scenario 1,  $p = 0$ ; that is, no sample information is available. For this scenario, even if the decision-maker knows the rewards  $r_i$  ( $i \in [2]$ ), without any sample information, as shown in Ball and Queyranne (2009),  $1/(2 - \alpha)$  is an upper bound for any algorithm.

Now consider scenario 2, where  $p$  is completely unknown. Suppose that the decision-maker has access to some sample information  $(s_1, s_2)$ , which means that there are  $s_1$  type 1 agents and  $s_2$  type 2 agents in the sample. As  $p$  is completely unknown,  $p$  can be any real number between 0 and 1. If  $p \rightarrow 0$ , the number of arriving agents of type 1  $n_1 \rightarrow \infty$  and  $n_2 \rightarrow \infty$ . If  $p = 1$ , we have  $n_1 = n_2 = 0$ . Therefore, the decision-maker has no demand information because  $n_i$  can be any non-negative real number for  $i \in \{1, 2\}$ . Thus, again by Ball and Queyranne (2009), the upper bound of CR among all deterministic and randomized algorithms is  $1/(2 - \alpha)$ .

## Appendix B: Proof of Statements in Section 4.1

### B.1. Proof of Lemma 1

We will show that for any realization of  $(n_1, n_2)$  and sample information  $\psi$ , we have  $\text{REW}_A(I_{\text{order}}, \psi) \leq \text{REW}_A(I, \psi)$ . This implies that  $\frac{\text{REW}_A(I_{\text{order}}, \psi)}{\text{OPT}(I_{\text{order}})} \leq \frac{\text{REW}_A(I, \psi)}{\text{OPT}(I)}$ , as  $\text{OPT}(I) = \text{OPT}(I_{\text{order}})$ . The last inequality is the desired result.

It remains to show that  $\text{REW}_A(I_{\text{order}}, \psi) \leq \text{REW}_A(I, \psi)$ . First consider a case where under  $\psi$ ,  $\hat{r}_1 > \hat{r}_2$ . In this case, Algorithm 1 assigns a protection level of  $x_1 = s_1(1 - p)/p$  to type 1 agents. Then, it is clear that (i) the number of accepted type 2 agents under  $I_{\text{order}}$  is larger than or equal to that under  $I$ , and (ii) the number of accepted type 1 agents under  $I_{\text{order}}$  is smaller than or equal to that under  $I$ . Hence in this case, Algorithm 1 under  $I_{\text{order}}$  obtains a lower reward than  $I$ . On the other hand, when  $\hat{r}_1 \leq \hat{r}_2$ , Algorithm 1 assigns a protection level of  $x_2 = s_2(1 - p)/p$  to type 2 agents. Again, in this case, we have (i) the number of accepted type 2 agents under  $I_{\text{order}}$  is larger than or equal to that under  $I$ , and (ii) the number of accepted type 1 agents under  $I_{\text{order}}$  is smaller than or equal to that under  $I$ . Hence, we have  $\text{REW}_A(I_{\text{order}}, \psi) \leq \text{REW}_A(I, \psi)$ .

### B.2. Proof of Lemma 2

By Equation (7),  $\inf_{(h, \ell) \in \mathcal{R}_1} \mathbb{E}_\psi \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$  can be lower bounded as follows:

$$\inf_{(h, \ell) \in \mathcal{R}_1} \mathbb{E}_\psi \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] \geq \inf_{(h, \ell) \in \mathcal{R}_1} \left\{ \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right], \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})} \right] \right\} \right\}. \quad (9)$$

In light of Equation (9), we divide the rest of the proof into two parts where in the first part, for any realization of  $\mathbf{s}$ , we bound  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})}$  and in the second part, we bound  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})}$ .

*Part 1: bounding  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})}$ .* Recall that for a fixed realization of  $\mathbf{s}$  (or equivalently  $\mathbf{n}$ ),  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})$  is the reward of our algorithm when there are  $n_i$  type  $i \in \{1, 2\}$  agents in the online arrival sequence given

that  $\hat{r}_1 > \hat{r}_2$ . (See the definition of the good event  $G(\boldsymbol{\psi}; (h, \ell))$  in Equation (4)). Conditioned on  $\hat{r}_1 > \hat{r}_2$ , Algorithm 1 assigns a protection level of  $x_1 = \min\{m, \frac{1-p}{p}s_1\}$  to type 1 agents, and hence we have

$$\begin{aligned} \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) &= \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\} \cdot r_2 \\ &\quad + \min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}, n_1 \right\} \cdot r_1, \end{aligned} \quad (10)$$

where  $\min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}$  is the number of accepted agents of type 2 and  $\min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}, n_1 \right\}$  is the number of accepted agents of type 1. In what follows, we bound  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})}$  by considering the following cases:

- *Case 1:*  $n_2 < m - n_1$ . In this case, under the optimal clairvoyant algorithm, all agents are accepted as we have  $\min\{(m - n_1)^+, n_2\} = n_2$ , and hence  $\text{OPT}(\mathbf{n}) = n_2r_2 + n_1r_1$ . In addition, the number of accepted type 1 agents under Algorithm 1, i.e.,  $\min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}, n_1 \right\}$ , is  $n_1$ ; see Equation (10). To see why note that  $\min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\} \leq n_2$ , and we have  $n_1 < m - n_2 \leq m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}$ , which implies that  $\min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}, n_1 \right\} = n_1$ . Then, by Equation (10), we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &= \frac{\min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\} \cdot r_2 + n_1r_1}{n_2r_2 + n_1r_1} \\ &\geq \frac{\min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}}{n_2} = \min \left\{ 1, \frac{\left( m - \frac{1-p}{p}s_1 \right)^+}{n_2} \right\} \\ &= \frac{\left( m - \frac{1-p}{p}s_1 \right)^+}{n_2} > \frac{\left( m - \frac{1-p}{p}s_1 \right)^+}{m - n_1} \geq \frac{\left( m - \frac{1-p}{p}h \right)^+}{m} \geq \left( 1 - \frac{1-p}{p} \frac{1}{\sqrt{m}} \right)^+, \end{aligned}$$

where in the last inequality holds because in region  $\mathcal{R}_1$ ,  $h < \sqrt{m}$ .

- *Case 2:*  $n_2 \geq m - n_1$ . In this case, the optimal clairvoyant algorithm, cannot accept all the agents in the online arrival sequence as we have  $\min\{(m - n_1)^+, n_2\} = m - n_1$ . In addition, since  $n_1 \leq h < \sqrt{m} < m$ , we have  $\text{OPT}(\mathbf{n}) = n_1r_1 + (m - n_1)r_2$ . Next, to bound  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})}$ , we consider the following two subcases:
  - *Case 2.1:*  $\min\{m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\}, n_1\} = m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\}$ . In this case, Algorithm 1 does not accept all type 1 agents (see Equation (10)): the number of accepted type 1 agents is  $m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\}$ . Then, by Equation (10), we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &= \frac{\min\{n_2, (m - \frac{1-p}{p}s_1)^+\}r_2 + (m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\})r_1}{(m - n_1)r_2 + n_1r_1} \\ &\geq \frac{mr_2}{(m - h)r_2 + hr_1} \geq \frac{mr_2}{(m - \sqrt{m})r_2 + \sqrt{m}r_1} = 1 - \frac{\sqrt{m}r_1 - \sqrt{m}r_2}{\sqrt{m}r_1 + (m - \sqrt{m})r_2}. \end{aligned}$$

The first inequality is due to  $\min\{n_2, (m - \frac{1-p}{p}s_1)^+\}r_2 + (m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\})r_1 \geq \min\{n_2, (m - \frac{1-p}{p}s_1)^+\}r_2 + (m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\})r_2 = mr_2$ .

— *Case 2.2:*  $\min\{\frac{1-p}{p}s_1, n_1\} = n_1$ . In this case, Algorithm 1 accepts all type 1 agents. We further have  $n_2 \geq m - n_1 \geq (m - \frac{1-p}{p}s_1)^+$ . Then, by Equation (10), we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &= \frac{\min\{n_2, (m - \frac{1-p}{p}s_1)^+\}r_2 + n_1r_1}{(m - n_1)r_2 + n_1r_1} \\ &= \frac{(m - \frac{1-p}{p}s_1)^+r_2 + n_1r_1}{(m - n_1)r_2 + n_1r_1} \geq \frac{(m - \frac{1-p}{p}s_1)^+}{m - n_1} \geq \frac{(m - \frac{1-p}{p}h)^+}{m} \geq \left(1 - \frac{1-p}{p} \frac{1}{\sqrt{m}}\right)^+. \end{aligned}$$

Putting these cases together, we have

$$\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \geq \min \left\{ \left(1 - \frac{1-p}{p} \frac{1}{\sqrt{m}}\right)^+, 1 - \frac{\sqrt{m}r_1 - \sqrt{m}r_2}{\sqrt{m}r_1 + (m - \sqrt{m})r_2} \right\}. \quad (11)$$

*Part 2: bounding  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})}$ .* Recall that for fixed realization of  $\mathbf{s}$ ,  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)$  is the reward of our algorithm when there are  $n_i$  type  $i \in \{1, 2\}$  agents in the online arrival sequence and  $\hat{r}_1 \leq \hat{r}_2$ . In this case, the algorithm assigns a protection level of  $x_2 = \frac{1-p}{p}s_2$ , and hence we have

$$\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C) \geq \min\{m, n_2\} \cdot r_2 + \min\{n_1, (m - n_2)^+\} \cdot r_1. \quad (12)$$

To bound our desired ratio, we consider the following cases:

- *Case 1:*  $n_2 \leq m - n_1$ . In this case, we have  $\min\{m, n_2\} = n_2$ , and  $\min\{n_1, (m - n_2)^+\} = n_1$ , and hence, by Equation (12),  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C) \geq n_1 \cdot r_1 + n_2 \cdot r_2$ . Thus, we have

$$\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \geq \frac{n_1r_1 + n_2r_2}{n_1r_1 + n_2r_2} = 1.$$

- *Case 2:*  $n_2 > m - n_1$ . In this case, we have  $\min\{n_1, (m - n_2)^+\} = (m - n_2)^+$ ,  $\min\{(m - n_1)^+, n_2\} = m - n_1$ . Then, by Equation (12), we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} &\geq \frac{\min\{m, n_2\}r_2 + (m - n_2)^+r_1}{n_1r_1 + (m - n_1)r_2} \\ &\geq \frac{mr_2}{\sqrt{m}r_1 + (m - \sqrt{m})r_2} = 1 - \frac{\sqrt{m}r_1 - \sqrt{m}r_2}{\sqrt{m}r_1 + (m - \sqrt{m})r_2}. \end{aligned}$$

Considering the following two cases, we have

$$\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \geq 1 - \frac{\sqrt{m}r_1 - \sqrt{m}r_2}{\sqrt{m}r_1 + (m - \sqrt{m})r_2}. \quad (13)$$

By plugging in Equations (11) and (13) into Equation (9), we get the desired result.

### B.3. Proof of Lemma 3

We follow similar steps in the proof of Lemma 2. That is, we bound

$$\inf_{(h, \ell) \in \mathcal{R}_2} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] \geq \inf_{(h, \ell) \in \mathcal{R}_2} \left\{ \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right], \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \right] \right\} \right\}. \quad (14)$$

In light of Equation (14), we divide the rest of the proof into two parts where in the first part, for any realization of  $\mathbf{s}$ , we bound  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})}$  and in the second part, we bound  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$ . While bounding the former ratio is straightforward, bounding the latter one is quite involved. This is because when the

good event  $\mathcal{G}(\boldsymbol{\psi}; (h, \ell))$  happens, Algorithm 1 may still over-protect or under-protect type 1 agents given the randomness in the sample information. As  $h$  is large in this case, the number of type 1 agents that the algorithm over-protects or under-protects can be large too, leading to a low reward for the algorithm under some realization of  $\mathbf{s}$ . Nonetheless, we show that the expected reward of the algorithm, in this case, is still large once we take expectation with respect to  $\mathbf{s}$ .

*Part 1: bounding  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})}$ .* Recall that for fixed realization of  $\mathbf{s}$ ,  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)$  is the reward of our algorithm when there are  $n_i$  type  $i \in \{1, 2\}$  agents in the online arrival sequence and  $\hat{r}_1 \leq \hat{r}_2$ . In this case, the algorithm assigns a protection level of  $x_2 = \min\{m, \frac{1-p}{p}s_2\}$ , and hence we have

$$\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C) \geq \min\{m, n_2\}r_2 + \min\{n_1, (m - n_2)^+\}r_1.$$

As  $\ell < \sqrt{m}$ , we have  $n_2 \leq \ell < \sqrt{m} < m$ . Hence,

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} &\geq \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{n_1r_1 + \min\{m - n_1, n_2\}r_2} \\ &= \frac{n_2r_2 + \min\{m - n_2, n_1\}r_1}{n_1r_1 + \min\{m - n_1, n_2\}r_2} \\ &\geq \min\left\{\frac{n_1r_1 + n_2r_2}{n_1r_1 + n_2r_2}, \frac{(m - n_2)r_1 + n_2r_2}{n_1r_1 + (m - n_1)r_2}\right\} \\ &\stackrel{a}{=} \frac{(m - n_2)r_1 + n_2r_2}{n_1r_1 + (m - n_1)r_2} \geq \frac{(m - \sqrt{m})r_1}{mr_1} = 1 - \frac{1}{\sqrt{m}}. \end{aligned}$$

The second inequality is because when  $\min\{m - n_2, n_1\} = n_1$ , we have  $n_2 < m - n_1$  and  $\min\{m - n_1, n_2\} = n_2$ , and when  $\min\{m - n_2, n_1\} = m - n_2$ , we have  $m - n_1 < n_2$  and  $\min\{m - n_1, n_2\} = m - n_1$ .

*Part 2: bounding  $\inf_{(h, \ell) \in \mathcal{R}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$ .* We begin by partitioning region  $\mathcal{R}_2$  based on the number of type 1 agents, (i.e.  $h$ ):

$$\overline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h \geq h_0\} \quad \underline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h < h_0\},$$

where  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . Then,  $\inf_{(h, \ell) \in \mathcal{R}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$  is equal to the minimum of  $\inf_{(h, \ell) \in \underline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$  and  $\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$ . We bound the first ratio (i.e., the one concerns  $\underline{\mathcal{R}}_2$ ) in Lemma 5 and the second ratio (i.e., the one concerns  $\overline{\mathcal{R}}_2$ ) will be bounded in Lemma 6. In addition,  $h_0 > \sqrt{m}$  is guaranteed because  $h_0 = \frac{m - \sqrt{m}}{1-p} + \frac{1}{2p^2} + \sqrt{\frac{1}{4p^4} + \frac{(m - \sqrt{m})}{p^2(1-p)}} > m - \sqrt{m} + \sqrt{2(m - \sqrt{m})} > \sqrt{m}$  for all  $m > 0$ .

**LEMMA 5.** *Let  $\underline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h < h_0\}$ , where  $\mathcal{R}_2 = \{(h, \ell) : h \geq \sqrt{m}, \ell < \sqrt{m}\}$ ,  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . Then,*

$$\inf_{(h, \ell) \in \underline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \geq \left(1 - \frac{1}{\sqrt{m}}\right) \left(1 - \frac{1}{m}\right).$$

Lemma 5 gives the lower bound of the ratio conditional on the good event happens. In this case, the algorithm gives a certain protection level to type 1 agents. Then, we show that when  $(h, \ell) \in \underline{\mathcal{R}}_2$ , with high probability, the number of type 2 agents is less than the total number of resources minus the protection level. Thus, in this regime, we do not reject any type 2 agents with high probability. Conditional on this event, we can obtain a lower bound of the ratio for any realization  $\mathbf{s}$ .

Next, we lower bound  $\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$  in Lemma 6.

LEMMA 6. Let  $\overline{\mathcal{R}}_2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_2, h \geq h_0\}$ , where  $\mathcal{R}_2 = \{(h, \ell) : h \geq \sqrt{m}, \ell < \sqrt{m}\}$ , where  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ ,  $h_0 = \min\{y \geq 0 : \frac{p(m-\sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . Then,

$$\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \geq \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\}.$$

To show Lemma 6, notice that given the good event happens, the algorithm may over- or under-protect type 1 agents. If it under-protects type 1 agents, we further split the analysis into two cases: the number of type 2 agents is less than, or is larger than the total number of resources minus the protection level. In both cases, we can easily find a lower bound for any realization  $\mathbf{s}$ . If the algorithm over-protects type 1 agents, then it accepts all arriving type 1 agents, but some resources are wasted due to over-protecting. In this case, we lower bound the ratio by a linear function with respect to the number of arriving type 1 agents,  $n_1$ . Then, we used the results in Nagaev and Chebotarev (2011) and Berend and Kontorovich (2013) to estimate the conditional expectation of  $n_1$  given that the algorithm over-protects type 1 agents.

Finally, by Lemmas 5 and 6, we have the following inequality, which is the desired result.

$$\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{s}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \geq \text{CR}_2 = \min \left\{ \left(1 - \frac{1}{\sqrt{m}}\right) \left(1 - \frac{1}{m}\right), \frac{h_1}{m} \right\}.$$

**B.3.1. Proof of Lemma 5** Define event  $\mathcal{E}_0$  as  $\{n_2 \leq m - \frac{1-p}{p}s_1\}$ . We provide a lower bound on  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right]$  as follows:

$$\begin{aligned} & \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \\ &= \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_0^c \right] \Pr(\mathcal{E}_0^c) + \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_0 \right] \Pr(\mathcal{E}_0) \\ &\geq \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_0 \right] \Pr(\mathcal{E}_0) \end{aligned} \quad (15)$$

In light of Equation (15), we will provide a lower bound for  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_0 \right]$  and  $\Pr(\mathcal{E}_0)$ .

*Part 1: bounding  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_0 \right]$ .* For a fixed realization of  $\mathbf{s}$  (or equivalently  $\mathbf{n}$ ), we have

$$\begin{aligned} \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) &= \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\} \cdot r_2 \\ &\quad + \min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p}s_1 \right)^+ \right\}, n_1 \right\} \cdot r_1 \end{aligned} \quad (16)$$

Conditional on  $\mathcal{E}_0 = \{n_2 \leq m - \frac{1-p}{p}s_1\}$ , we have  $\min\{n_2, (m - \frac{1-p}{p}s_1)^+\} = n_2$ , and  $\min\{m - \min\{n_2, (m - \frac{1-p}{p}s_1)^+\}, n_1\} = \min\{m - n_2, n_1\}$ . Then, we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &\geq \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{n_1 r_1 + \min\{m - n_1, n_2\} r_2} \\ &= \frac{n_2 r_2 + \min\{m - n_2, n_1\} r_1}{n_1 r_1 + \min\{m - n_1, n_2\} r_2} \\ &\geq \min \left\{ \frac{n_1 r_1 + n_2 r_2}{n_1 r_1 + n_2 r_2}, \frac{(m - n_2) r_1 + n_2 r_2}{n_1 r_1 + (m - n_1) r_2} \right\} \\ &= \frac{(m - n_2) r_1 + n_2 r_2}{n_1 r_1 + (m - n_1) r_2} \geq \frac{(m - \sqrt{m}) r_1}{m r_1} = 1 - \frac{1}{\sqrt{m}}. \end{aligned} \quad (17)$$

The second inequality is because when  $\min\{m - n_2, n_1\} = n_1$ , we have  $\min\{m - n_1, n_2\} = n_2$ , and when  $\min\{m - n_2, n_1\} = m - n_2$ , we have  $\min\{m - n_1, n_2\} = m - n_1$ .

*Part 2: bounding  $\Pr(\mathcal{E}_0)$ .* We want to show that  $\Pr(\mathcal{E}_0)$  is a high probability event for  $(h, \ell) \in \underline{\mathcal{R}}_2$ . As  $\ell \in \sqrt{m}$  for any  $(h, \ell) \in \underline{\mathcal{R}}_2$ , we have  $n_2 \leq \sqrt{m}$ . Then, because  $\mathcal{E}_0 = \{n_2 \leq m - \frac{1-p}{p}s_1\}$ , we have

$$\Pr(\mathcal{E}_0) = \Pr\left(\frac{1-p}{p}s_1 \leq m - n_2\right) \geq \Pr\left(\frac{1-p}{p}s_1 \leq m - \sqrt{m}\right) = \Pr\left(s_1 \leq \frac{p(m - \sqrt{m})}{1-p}\right).$$

Next, we show that with high probability, the upper bound of  $s_1$  is no more than  $\frac{p(m - \sqrt{m})}{1-p}$ . Because  $s_1 \sim \text{Bin}(h, p)$ , by Hoeffding's inequality, we have

$$\Pr(|s_1 - hp| \leq \sqrt{h}) \geq 1 - \frac{1}{h^2} \geq 1 - \frac{1}{m}.$$

This implies that  $s_1 \leq hp + \sqrt{h}$  happens with high probability. We then have

$$1 - \frac{1}{m} \leq \Pr(s_1 \leq hp + \sqrt{h}) \leq \Pr(s_1 \leq h_0p + \sqrt{h_0}) \leq \Pr\left(s_1 \leq \frac{p(m - \sqrt{m})}{1-p}\right), \quad (18)$$

where the second inequality holds because  $hp + \sqrt{h}$  is an increasing function, and  $h < h_0$  in region  $\underline{\mathcal{R}}_2$ . These imply that  $hp + \sqrt{p} \leq h_0p + \sqrt{p_0}$ . The last inequality holds because  $\frac{p(m - \sqrt{m})}{1-p} \geq h_0p + \sqrt{h_0}$ . Recall that  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ .

By Equations (17) and (18), we have

$$\inf_{(h, \ell) \in \underline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_0 \right] \Pr(\mathcal{E}_0) \geq \left(1 - \frac{1}{\sqrt{m}}\right) \left(1 - \frac{1}{m}\right), \quad (19)$$

which is the desired result.

**B.3.2. Proof of Lemma 6** We define event  $\mathcal{E}_1 = \{n_1 > \frac{1-p}{p}s_1\}$ . We then have

$$\begin{aligned} \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] &\geq \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \left\{ \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1 \right], \right. \right. \\ &\quad \left. \left. \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1^c \right] \right\} \right\} \end{aligned} \quad (20)$$

We will provide the lower bound of  $\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1 \right]$  in the first part of the proof, and the lower bound of  $\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1^c \right]$  in the second part of the proof.

*Part 1: bounding  $\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1 \right]$ .* We then have

$$\begin{aligned} &\inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1 \right] \\ &\geq \inf_{(h, \ell) \in \overline{\mathcal{R}}_2} \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0^c \right], \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0 \right] \right\}, \end{aligned}$$

where we recall  $\mathcal{E}_0^c = \{n_2 > m - \frac{1-p}{p}s_1\}$ .

To provide a lower bound for  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0^c \right]$ , we note that under event  $\mathcal{E}_0^c$ , we have  $n_2 > m - \frac{1-p}{p}s_1$ , and hence

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0^c \right] \geq \frac{(m - \frac{1-p}{p}s_1)^+ r_2 + \min\{m, \frac{1-p}{p}s_1\} r_1}{mr_1} \geq \frac{(m - n_2)r_1}{mr_1} \geq 1 - \frac{1}{\sqrt{m}}. \quad (21)$$

Then, we give a lower bound to  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0 \right]$ . Conditional on  $\mathcal{E}_1$  and  $\mathcal{E}_0$ , we have  $\frac{1-p}{p} s_1 \geq m - n_2$  and  $\frac{1-p}{p} s_1 \leq n_1$ . That is, the number of type 1 agents in the online arrival sequence is greater than or equal to the protection level of Algorithm 1 for type 1 agents (i.e.,  $\frac{1-p}{p} s_1 \leq n_1$ ), and in addition, the algorithm ends up rejecting some of type 2 agents as  $\frac{1-p}{p} s_1 \geq m - n_2$ . Therefore, we have

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1, \mathcal{E}_0 \right] \geq \frac{(m - \frac{1-p}{p} s_1)^+ r_2 + \min\{m, \frac{1-p}{p} s_1\} r_1}{mr_1} \geq \frac{(m - n_2) r_1}{mr_1} \geq 1 - \frac{1}{\sqrt{m}},$$

where the last inequality holds because  $n_2 \leq \ell \leq \sqrt{m}$  for any  $(h, \ell) \in \mathcal{R}_2$ .

*Part 2: bounding  $\inf_{(h, \ell) \in \mathcal{R}_2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1^c \right]$ .* Recall that for a fixed realization of  $\mathbf{s}$  (or equivalently  $\mathbf{n}$ ),

$$\begin{aligned} \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) &= \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\} \cdot r_2 \\ &\quad + \min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\}, n_1 \right\} \cdot r_1. \end{aligned}$$

Conditional on  $\mathcal{E}_1^c = \{n_1 \leq \frac{1-p}{p} s_1\}$ , we do not receive enough number of type 1 agents and thus, we have

$$\min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\}, n_1 \right\} = \min \left\{ \max \left\{ m - n_2, \min\{m, \frac{1-p}{p} s_1\} \right\}, n_1 \right\} = n_1.$$

That is, Algorithm 1 accepts all type 1 agents in the online arrival sequence. Then, we have

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_1^c \right] \geq \mathbb{E}_{\mathbf{n}} \left[ \frac{\min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\} \cdot r_2 + n_1 r_1}{mr_1} \mid \mathcal{E}_1^c \right] \geq \frac{1}{m} \mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^c].$$

We will show that  $\mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^c] \geq h_1$ , where  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$  and  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ .

Since  $\mathcal{E}_1^c = \{n_1 \leq \frac{1-p}{p} s_1\} = \{h(1-p) - n_1 \geq 0\}$ , and  $h(1-p) = \mathbb{E}[h(1-p)]$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^c] &= h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid h(1-p) - n_1 \geq 0] \\ &\geq h(1-p) - \mathbb{E}_{\mathbf{s}} [|h(1-p) - n_1|] - \frac{\beta}{\sqrt{h}} \\ &\geq h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}} \\ &\geq h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}} \\ &= h_1, \end{aligned}$$

where the first inequality holds because by Nagaev and Chebotarev (2011), we have  $\left| \mathbb{E}_{\mathbf{s}} \left[ (h(1-p) - n_1) \mid \{h(1-p) - n_1 \geq 0\} \right] - \mathbb{E}_{\mathbf{s}} [|h(1-p) - n_1|] \right| \leq \frac{\beta}{\sqrt{h}}$ . The second inequality follows from Berend and Kontorovich (2013) that shows  $\mathbb{E}_{\mathbf{s}} [|h(1-p) - n_1|] \leq \sqrt{p(1-p)h}$ . The third inequality holds because when  $h > h_0$ ,  $h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}}$  is a non-decreasing function for any  $p$ . To see why note that

$$\frac{\partial (h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}})}{\partial h} \geq (1-p) - \frac{1}{2} \sqrt{p(1-p)} \frac{1}{\sqrt{h}} \geq 0. \quad (22)$$

where the last inequality holds for any  $h \geq \frac{1-p}{4}$ , and we have  $h_0 > \frac{m-\sqrt{m}}{1-p} > \frac{1-p}{4}$  for all  $m \geq 2$ .



#### B.4. Proof of Lemma 4

By Equation (7), we have

$$\begin{aligned} \inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\psi} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] &= \inf_{(h,\ell) \in \mathcal{R}_3} \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c) \cdot \Pr(\mathcal{G}^c(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right\} \end{aligned} \quad (23)$$

We split the analysis into two cases based on the initial number of resources  $m$ . In the first case, we assume that  $m \geq m_1$ , and in the second case, we assume that  $m < m_1$ . Here,  $m_1 = \min_{y \geq 1/p^4} \left\{ y : r_1 - r_2 > \frac{2}{y^{1/8} \sqrt{y^{1/4} p - 1}} \right\}$  is a constant that only depends on  $p$ ,  $r_1$ , and  $r_2$ . Case 1 is considered in Lemma 7 while case 2 is studied in Lemma 8.

LEMMA 7. *Let  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ . Let  $m_1 = \min_{y \geq 1/p^4} \left\{ y : r_1 - r_2 > \frac{2}{y^{1/8} \sqrt{y^{1/4} p - 1}} \right\}$ .*

*When the initial number of resources  $m \geq m_1$ , we have*

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] \geq \left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2 W.$$

where  $W = \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\}$ ,  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ , and  $h_0 = \min \{ y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y} \}$ .

To show Lemma 7, we observe that when the initial number of resources  $m \geq m_1$ , the good event happens with high probability. Then, conditional on the algorithm protects type 1 agents, we give the lower bound of the ratio by the similar methodology introduced in Lemma 5 and Lemma 6: we discuss whether the algorithm over- or under-protects type 1 agents, and whether the number of type 2 agents in the arriving sequence is less than, or is larger than the total number of resources minus the protection level.

LEMMA 8. *Let  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ ,  $m_1 = \min_{y \geq 1/p^4} \left\{ y : r_1 - r_2 > \frac{2}{y^{1/8} \sqrt{y^{1/4} p - 1}} \right\}$ , and  $\alpha = \frac{r_2}{r_1}$ . When the initial number of resources  $m < m_1$ , we have*

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\mathbb{E}[\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})]}{\text{OPT}(\mathbf{n})} \right] \geq V \cdot \min \left\{ \left(1 - \frac{1}{m^2}\right)W, \frac{1}{2} \left(1 - \frac{1}{m^2}\right)W + \frac{1}{2} \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right)\alpha, 1 - \frac{1-p}{pm} \ell_1 \right\} \right\}.$$

where  $V = 1 - 2(1-p)\sqrt{m}$ ,  $W = \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\}$ ,  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ ,  $h_0 = \min \{ y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y} \}$ ,  $\ell_0 = \min \{ y : (1-p)y - \sqrt{y} \geq m \}$ , and  $\ell_1 = \sqrt{p(1-p)\ell_0} + \frac{\beta}{\sqrt{m}} + \ell_0 p$ .

To show Lemma 8, we should use a different technique rather than the one we use in Lemma 7. This is because if the initial number of resources  $m < m_1$ , the bad event is no longer a rare event. Then, if the good event happens, the ratio is derived in Lemma 7. If the bad event happens, the algorithm gives protection level to type 2 agents. Then, we first consider that if the algorithm under-protects type 2 agents, the algorithm accepts all type 2 agents, and give the remaining resources to type 1 agents. In this case, the ratio is bounded by  $\alpha$ . Otherwise, if the algorithm over-protects type 2 agents, the algorithm rejects some type 1 agents and wastes some resources. Then, we use concentration inequality to bound the loss.

**B.4.1. Proof of Lemma 7** We start by defining the event

$$\mathcal{E}_2(h, \ell) = \left\{ s_1 \in [hp - \sqrt{h}, hp + \sqrt{h}] \cap \{s_2 \in [\ell p - \sqrt{\ell}, \ell p + \sqrt{\ell}] \right\}.$$

Under event  $\mathcal{E}_2(h, \ell)$ , both  $s_1$  and  $s_2$  are concentrated around their average. Recall that  $s_1 \sim \text{Bin}(h, p)$  and  $s_2 \sim \text{Bin}(\ell, p)$ . By Equation (7), we have

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_\psi \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] \quad (24)$$

$$\geq \inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \Big| \mathcal{E}_2(h, \ell) \right] \Pr(\mathcal{E}_2(h, \ell)). \quad (25)$$

We lower bound  $\Pr(\mathcal{E}_2(h, \ell))$  in Part 1 of the proof. Next, we explain how to handle the conditional expectation in Equation (25). We start by defining the following event

$$\mathcal{E}_3 = \left\{ \hat{r}_1 \in \left[ r_1 - \frac{1}{\sqrt{s_1}}, r_1 + \frac{1}{\sqrt{s_1}} \right] \cap \{ \hat{r}_2 \in \left[ r_2 - \frac{1}{\sqrt{s_2}}, r_2 + \frac{1}{\sqrt{s_2}} \right] \right\}.$$

Under event  $\mathcal{E}_3$ , our estimates for  $r_1$  and  $r_2$  are concentrated around their true values. The conditional expectation in Equation (25) is lower bounded as follows:

$$\begin{aligned} & \inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \Big| \mathcal{E}_2(h, \ell) \right] \\ & \geq \inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \Big| \mathcal{E}_2(h, \ell), \mathcal{E}_3 \right] \Pr(\mathcal{E}_3 | \mathcal{E}_2(h, \ell)) \end{aligned} \quad (26)$$

Given (26), we lower bound  $\Pr(\mathcal{E}_3 | \mathcal{E}_2(h, \ell))$  in part 2 of the proof. In part 3, conditioned on  $\mathcal{E}_2(h, \ell), \mathcal{E}_3$ , we show that  $\Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))$ , for any realization  $\mathbf{s}$ , equals to 1 in part 3. Finally, in part 4, we we lower bound  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \Big| \mathcal{E}_2(h, \ell), \mathcal{E}_3 \right]$ .

*Part 1: bounding  $\Pr(\mathcal{E}_2(h, \ell))$ .* By Hoeffding's inequality, we have

$$\Pr(s_1 \in [hp - \sqrt{h}, hp + \sqrt{h}]) \geq 1 - \frac{1}{h^2} \geq 1 - \frac{1}{m},$$

$$\Pr(s_2 \in [\ell p - \sqrt{\ell}, \ell p + \sqrt{\ell}]) \geq 1 - \frac{1}{\ell^2} \geq 1 - \frac{1}{m}.$$

The last steps of both equations are because in region  $\mathcal{R}_3$ , we have  $h \geq \sqrt{m}$ ,  $\ell \geq \sqrt{m}$ . Then, because  $s_1$  and  $s_2$  are independent of each other, we have

$$\Pr(\mathcal{E}_2(h, \ell)) = \Pr\left(s_1 \in [hp - \sqrt{h}, hp + \sqrt{h}]\right) \cdot \Pr\left(s_2 \in [\ell p - \sqrt{\ell}, \ell p + \sqrt{\ell}]\right) \geq \left(1 - \frac{1}{m}\right)^2.$$

*Part 2: bounding  $\Pr(\mathcal{E}_3 | \mathcal{E}_2(h, \ell))$ .* Conditioned on  $s_1$ , we know  $\hat{r}_1 = \frac{\text{Bin}(s_1, r_1)}{s_1}$ , and similarly conditioned on  $s_2$ ,  $\hat{r}_2 = \frac{\text{Bin}(s_2, r_2)}{s_2}$ . Then, Hoeffding's inequality implies that

$$\Pr\left(\hat{r}_1 \in \left[ r_1 - \frac{1}{\sqrt{s_1}}, r_1 + \frac{1}{\sqrt{s_1}} \right] \mid s_1 = \tilde{s}_1\right) \geq 1 - \frac{1}{\tilde{s}_1^2}. \quad (27)$$

Now consider event  $\mathcal{E}_2(h, \ell)$  that implies that  $s_1 \geq hp - \sqrt{h}$ . As shown in in the proof of Lemma 5,  $hp - \sqrt{h}$  is an increasing function for any  $h \geq \sqrt{m}$ , and hence under event  $\mathcal{E}_2(h, \ell)$ ,  $s_1 \geq hp - \sqrt{h} \geq \sqrt{mp} - m^{1/4}$ .

Applying this lower bound on  $s_1$  to Equation (27) leads to

$$\Pr\left(\hat{r}_1 \in \left[ r_1 - \frac{1}{\sqrt{s_1}}, r_1 + \frac{1}{\sqrt{s_1}} \right] \mid \mathcal{E}_2(h, \ell)\right) \geq 1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}.$$

Note that the same inequality holds  $r_2$  and its estimate. As a result, we have

$$\Pr(\mathcal{E}_3 | \mathcal{E}_2(h, \ell)) \geq \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2.$$

*Part 3: bounding  $\Pr(\mathcal{G}(\psi; (h, \ell)))$  conditioned on events  $\mathcal{E}_2(h, \ell)$  and  $\mathcal{E}_3$ .* We show that conditioned on  $\mathcal{E}_2(h, \ell)$  and  $\mathcal{E}_3$ , for any realization  $\mathbf{s}$ ,  $\Pr(\mathcal{G}(\psi; (h, \ell))) = 1$ . We show this by proving  $\hat{r}_1 \geq \hat{r}_2$  always happens under events  $\mathcal{E}_2(h, \ell)$  and  $\mathcal{E}_3$ . Under events  $\mathcal{E}_2(h, \ell)$  and  $\mathcal{E}_3$ , we know that (i)  $s_1, s_2 \geq \sqrt{mp} - m^{1/4}$  and (ii)  $\hat{r}_1 \in [r_1 - \frac{1}{\sqrt{s_1}}, r_1 + \frac{1}{\sqrt{s_1}}]$  and  $\hat{r}_2 \in [r_2 - \frac{1}{\sqrt{s_2}}, r_2 + \frac{1}{\sqrt{s_2}}]$ . To show the result, we confirm that the lower bound on  $\hat{r}_1$  is smaller than the upper bound on  $\hat{r}_2$ . That is,

$$r_2 + \frac{1}{\sqrt{s_2}} \leq r_1 - \frac{1}{\sqrt{s_1}} \quad \text{for any } s_1, s_2 \geq \sqrt{mp} - m^{1/4}$$

Given the lower bound on  $s_1$  and  $s_2$ , it suffices to show that

$$r_2 + \frac{1}{\sqrt{\sqrt{mp} - m^{1/4}}} \leq r_1 - \frac{1}{\sqrt{\sqrt{mp} - m^{1/4}}} \iff r_1 - r_2 \geq \frac{2}{m^{1/8} \sqrt{m^{1/4}p - 1}}$$

By definition of  $m_1$ , the above inequality holds for any  $m \geq m_1$ . Recall that  $m_1 = \min_{y \geq 1/p^4} \{y : r_1 - r_2 > \frac{2}{y^{1/8} \sqrt{y^{1/4}p - 1}}\}$ .

*Part 4: bounding  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_3 \right]$ .* First, observe that the ratio  $\frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})}$  is independent with  $\mathcal{E}_3$  because both the reward under the good event  $\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})$ , and the optimal reward, i.e.,  $\text{OPT}(\mathbf{n})$ , do not depend on  $\hat{r}_1$  and  $\hat{r}_2$ . We can simply delete  $\mathcal{E}_3$ , and bound  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right]$ .

We begin by partitioning region  $\mathcal{R}_3$  based the number of type 1 agents, (i.e..  $h$ ):

$$\mathcal{R}_3^1 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, h < h_0\} \quad \text{and} \quad \mathcal{R}_3^2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, h \geq h_0\},$$

where  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . Then,  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right]$  is equal to the minimum of  $\inf_{(h, \ell) \in \mathcal{R}_3^1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right]$  and  $\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right]$ .

LEMMA 9. Let  $\mathcal{R}_3^1 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, h < h_0\}$ . We then have

$$\inf_{(h, \ell) \in \mathcal{R}_3^1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right] \geq 1 - \frac{1}{\sqrt{m}},$$

where  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$  and event  $\mathcal{E}_2(h, \ell)$  is defined in Equation (24).

The proof of Lemma 9 is similar to the proof of Lemma 5, and is presented in Section B.4.3.

LEMMA 10. Let  $\mathcal{R}_3^2 = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, h \geq h_0\}$ . We then have

$$\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right] \geq \frac{h_1}{m},$$

where  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$ ,  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ , and  $h_0 = \min\{y \geq 0 : \frac{p(m - \sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . Further, event  $\mathcal{E}_2(h, \ell)$  is defined in Equation (24).

The proof of Lemma 10 is similar to the proof of Lemma 6 and for completeness, is presented in Section B.4.4. Lemmas 9 and 10 imply that

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \geq \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\} = W.$$

Finally, by the results in parts 1-4, we have

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \right] \geq \left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2 W,$$

which is the desired result.

**B.4.2. Proof of Lemma 8** In Equation (7), we have

$$\begin{aligned} \inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{s}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] &= \inf_{(h,\ell) \in \mathcal{R}_3} \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c) \cdot (1 - \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell))))}{\text{OPT}(\mathbf{n})} \right] \right\}. \end{aligned} \quad (28)$$

Define event  $\mathcal{E}_5 = \{s_1 \neq 0\} \cap \{s_2 \neq 0\}$ . We can lower bound Equation (28) by

$$\begin{aligned} \inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{s}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] &\geq \inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\boldsymbol{\psi}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \Pr(\mathcal{E}_5) \\ &\geq \inf_{(h,\ell) \in \mathcal{R}_3} \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) \cdot \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell)))}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c) \cdot (1 - \Pr(\mathcal{G}(\boldsymbol{\psi}; (h, \ell))))}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \right\} \Pr(\mathcal{E}_5). \end{aligned} \quad (29)$$

Note that

$$\begin{aligned} \Pr(\mathcal{E}_5) &= 1 - \Pr(\mathcal{E}_5^c) \geq 1 - \Pr(s_1 = 0) - \Pr(s_2 = 0) \\ &= 1 - \binom{h}{0} (1-p)^h - \binom{\ell}{0} (1-p)^\ell \geq 1 - 2 \binom{\sqrt{m}}{0} (1-p)^{\sqrt{m}} = 1 - 2(1-p)^{\sqrt{m}}. \end{aligned} \quad (30)$$

We denote  $V$  as  $1 - 2(1-p)^{\sqrt{m}}$ . Furthermore, conditioned on event  $\mathcal{E}_5$ , it is obvious that the probability of the good event  $\mathcal{G}(\boldsymbol{\psi}; (h, \ell)) = \{\hat{r}_1 \geq \hat{r}_2\}$  is greater than or equal 1/2. This is because when event  $\mathcal{E}_5$  holds, (i) our estimate for  $r_i$ ,  $i \in \{1, 2\}$  is simply the sample average of the  $s_i$  realized rewards, and (ii) we have  $r_1 > r_2$ . In light of Equation (29), if  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] < \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$ , then Equation (29) is lower bounded by setting the probability of the good event to be 1.

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{s}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]. \quad (31)$$

Otherwise, Equation (29) is lower bounded by setting the probability of good event to be  $\frac{1}{2}$ .

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{s}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \inf_{(h,\ell) \in \mathcal{R}_3} \frac{1}{2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] + \frac{1}{2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \quad (32)$$

Given Equations (31) and (32), we need to bound  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$  and  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$ . Notice that the former ratio can be lower bounded by

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] &\geq \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5, \mathcal{E}_2(h, \ell) \right] \Pr(\mathcal{E}_2(h, \ell)) \\ &= \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_2(h, \ell) \right] \Pr(\mathcal{E}_2(h, \ell)), \end{aligned}$$

where event  $\mathcal{E}_2(h, \ell) = \{s_1 \in [hp - \sqrt{h}, hp + \sqrt{h}] \cap \{s_2 \in [\ell p - \sqrt{\ell}, \ell p + \sqrt{\ell}]\}$ . The equality holds because  $\mathcal{E}_2(h, \ell) \subset \mathcal{E}_5$ . Then, by part 1 and 4 of the proof of Lemma 7, we have the former ratio is bounded by

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \left(1 - \frac{1}{m}\right)^2 W, \quad (33)$$

where  $W = \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\}$ .

Bounding the latter ratio, i.e.,  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$ , is quite involved because when the bad event  $\mathcal{G}^C(\boldsymbol{\psi}; (h, \ell))$  happens, Algorithm 1 may over-protect or under-protect type 2 agents given the randomness in the sample information. Over-protecting type 2 agents leads to reject a certain number of type 1 agents and waste some of the resources. As  $\ell$  is large in this case, the number of wasted resources can be large too. Under-protecting type 2 agents leads to accept many type 1 agents, which is a good thing for the algorithm. Nonetheless, we provide the lower bound of the ratio below and show that it is still not too small in this case once we take expectation with respect to  $\mathbf{s}$ .

We begin by partitioning region  $\mathcal{R}_3$  based the number of type 2 agents chosen by the adversary, (i.e.  $\ell$ ):

$$\overline{\mathcal{R}_3} = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell > \ell_0\} \quad \underline{\mathcal{R}_3} = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell \leq \ell_0\},$$

where  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$ . Then,  $\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$  is equal to the minimum of  $\inf_{(h, \ell) \in \overline{\mathcal{R}_3}} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$  and  $\inf_{(h, \ell) \in \underline{\mathcal{R}_3}} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$ . We bound the first ratio (i.e., the one concerns  $\overline{\mathcal{R}_3}$ ) in Lemma 11 and the second ratio (i.e., the one concerns  $\underline{\mathcal{R}_3}$ ) will be bounded in Lemma 12.

LEMMA 11. *Let  $\overline{\mathcal{R}_3} = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell > \ell_0\}$ , where  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ , and  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$ . Then,*

$$\inf_{(h, \ell) \in \overline{\mathcal{R}_3}} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \left(1 - \frac{1}{\ell_0^2}\right) \alpha,$$

where  $\alpha = \frac{r_2}{r_1}$ .

In this case, we show that with high probability, the number of type 2 agents in the online arrival sequence  $n_2 \geq m$ . Then, the algorithm accepts  $m$  type 2 agents, and the ratio is bounded by  $\alpha$ .

LEMMA 12. *Let  $\underline{\mathcal{R}_3} = \{(h, \ell) : (h, \ell) \in \mathcal{R}_3, \ell \leq \ell_0\}$ , where  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ , and  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$ . Let  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$  and  $\ell_1 = \sqrt{p(1-p)\ell_0} + \frac{\beta}{\sqrt{m}} + \ell_0 p$ . Then,*

$$\inf_{(h, \ell) \in \underline{\mathcal{R}_3}} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \min \left\{ \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\}.$$

To show Lemma 12, we define event  $\mathcal{E}_4 = \{n_2 \geq \frac{1-p}{p} s_2\}$ . Under this event the number of type 2 agents is greater than or equal to the protection level assigned to this type of agents. Then, we provide the lower bound of  $\inf_{(h, \ell) \in \underline{\mathcal{R}_3}} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right]$  by further conditioning on either  $\mathcal{E}_4$  or  $\mathcal{E}_4^C$ . For the part with  $\mathcal{E}_4$ , we do not reject type 1 agents unless there is no resource remaining. Therefore, the lower bound can be bounded by  $\alpha$ . For the part with  $\mathcal{E}_4^C$ , we directly calculate how many resources we waste, and then provide a lower bound.

Finally, combining the results from Lemmas 11 and 12, we have

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right) \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\}. \quad (34)$$

Combining the results from Equations (31), and (33), we have if  $(1 - \frac{1}{m})^2 W \leq \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right) \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\}$ ,

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_5 \right] \geq \left(1 - \frac{1}{m}\right)^2 W. \quad (35)$$

Otherwise, by Equations (32), and (34), we have

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_5 \right] \geq \frac{1}{2} \left(1 - \frac{1}{m}\right)^2 W + \frac{1}{2} \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right) \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\}. \quad (36)$$

Therefore, combining the results in Equations (29), (30), and (35), (36), we have

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho})}{\text{OPT}(\mathbf{n})} \right] \geq V \cdot \min \left\{ \left(1 - \frac{1}{m^2}\right) W, \frac{1}{2} \left(1 - \frac{1}{m^2}\right) W + \frac{1}{2} \min \left\{ \left(1 - \frac{1}{\ell_0^2}\right) \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\} \right\},$$

where recall that  $V = 1 - 2(1-p)^{\sqrt{m}}$ .

**B.4.3. Proof of Lemma 9** Recall that  $\mathcal{E}_0 = \{n_2 > m - \frac{1-p}{p} s_1\}$ . We provide a lower bound of  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_2(h, \ell) \right]$  by considering the conditional expectation on  $\mathcal{E}_0^C$ .

$$\begin{aligned} & \inf_{(h,\ell) \in \mathcal{R}_3^1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_2(h, \ell) \right] \\ & \geq \inf_{(h,\ell) \in \mathcal{R}_3^1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_2(h, \ell), \mathcal{E}_0 \right] \Pr(\mathcal{E}_0 | \mathcal{E}_2(h, \ell)). \end{aligned} \quad (37)$$

In light of Equation (37), we will provide a lower bound for  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_2(h, \ell), \mathcal{E}_0 \right]$  and  $\Pr(\mathcal{E}_0 | \mathcal{E}_2(h, \ell))$ .

*Part 1: bounding  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} \middle| \mathcal{E}_2(h, \ell), \mathcal{E}_0 \right]$ .* For a fixed realization of  $\mathbf{s}$  (or equivalently  $\mathbf{n}$ ), we have

$$\begin{aligned} \text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}) &= \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\} \cdot r_2 \\ &+ \min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\}, n_1 \right\} \cdot r_1 \end{aligned} \quad (38)$$

Conditional on  $\mathcal{E}_0 = \{n_2 \leq m - \frac{1-p}{p} s_1\}$ , we have  $\min\{n_2, (m - \frac{1-p}{p} s_1)^+\} = n_2$ , and  $\min\{m - \min\{n_2, (m - \frac{1-p}{p} s_1)^+\}, n_1\} = \min\{m - n_2, n_1\}$ . Then, we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &\geq \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \\ &= \frac{n_2 r_2 + \min\{m - n_2, n_1\} r_1}{n_1 r_1 + \min\{m - n_1, n_2\} r_2} \\ &\geq \min \left\{ \frac{n_1 r_1 + n_2 r_2}{n_1 r_1 + n_2 r_2}, \frac{(m - n_2) r_1 + n_2 r_2}{n_1 r_1 + (m - n_1) r_2} \right\} \\ &= \frac{(m - n_2) r_1 + n_2 r_2}{n_1 r_1 + (m - n_1) r_2} \geq \frac{(m - \sqrt{m}) r_1}{m r_1} = 1 - \frac{1}{\sqrt{m}}. \end{aligned} \quad (39)$$

The second inequality is because when  $\min\{m - n_2, n_1\} = n_1$ , we have  $\min\{m - n_1, n_2\} = n_2$ , and when  $\min\{m - n_2, n_1\} = m - n_2$ , we have  $\min\{m - n_1, n_2\} = m - n_1$ .

*Part 2: bounding  $\Pr(\mathcal{E}_0 | \mathcal{E}_2(h, \ell))$ .* We want to show that  $\Pr(\mathcal{E}_0 | \mathcal{E}_2(h, \ell))$  equals 1 for  $(h, \ell) \in \mathcal{R}_3^1$ . As  $\ell \in \sqrt{m}$  for any  $(h, \ell) \in \mathcal{R}_3^1$ , we have  $n_2 \leq \sqrt{m}$ . Then, because  $\mathcal{E}_0 = \{n_2 \leq m - \frac{1-p}{p} s_1\}$ , we have

$$\begin{aligned} \Pr(\mathcal{E}_0 | \mathcal{E}_2(h, \ell)) &= \Pr \left( \frac{1-p}{p} s_1 \leq m - n_2 \middle| \mathcal{E}_2(h, \ell) \right) \geq \Pr \left( \frac{1-p}{p} s_1 \leq m - \sqrt{m} \middle| \mathcal{E}_2(h, \ell) \right) \\ &= \Pr \left( s_1 \leq \frac{p(m - \sqrt{m})}{1-p} \middle| \mathcal{E}_2(h, \ell) \right). \end{aligned}$$

Next, we show that conditional on  $\mathcal{E}_2(h, \ell)$ , the upper bound of  $s_1$  is no more than  $\frac{p(m-\sqrt{m})}{1-p}$ . If  $\mathcal{E}_2(h, \ell)$  happens, we have  $s_1 \leq hp + \sqrt{h}$ . We then have

$$s_1 \leq hp + \sqrt{h} \Rightarrow s_1 \leq h_0p + \sqrt{h_0} \leq s_1 \Rightarrow \frac{p(m-\sqrt{m})}{1-p},$$

where the second inequality holds because  $hp + \sqrt{h}$  is an increasing function, and  $h < h_0$  in region  $\mathcal{R}_2$ . These imply that  $hp + \sqrt{p} \leq h_0p + \sqrt{p_0}$ . The last inequality holds because  $\frac{p(m-\sqrt{m})}{1-p} \geq h_0p + \sqrt{h_0}$ . Recall that  $h_0 = \min\{y \geq 0 : \frac{p(m-\sqrt{m})}{1-p} \geq yp + \sqrt{y}\}$ . This means that

$$\Pr(\mathcal{E}_0^c | \mathcal{E}_2(h, \ell)) = 1. \quad (40)$$

By Equations (39) and (40), we have

$$\inf_{(h, \ell) \in \mathcal{R}_3^1} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_0 \right] \Pr(\mathcal{E}_0) \geq \left(1 - \frac{1}{\sqrt{m}}\right), \quad (41)$$

**B.4.4. Proof of Lemma 10** Recall that  $\mathcal{E}_1 = \{n_1 > \frac{1-p}{p}s_1\}$ . We bound  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right]$  by considering the conditional expectation on  $\mathcal{E}_1$  and  $\mathcal{E}_1^c$  respectively.

$$\begin{aligned} & \inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell) \right] \\ & \geq \inf_{(h, \ell) \in \mathcal{R}_3^2} \left\{ \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1 \right], \right. \right. \\ & \quad \left. \left. \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1^c \right] \right\} \right\} \end{aligned} \quad (42)$$

We will provide the lower bound of  $\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1 \right]$  in the first part of the proof, and the lower bound of  $\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1^c \right]$  in the second part of the proof.

*Part 1: bounding  $\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1 \right]$ .* We then have

$$\begin{aligned} & \inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1 \right] \\ & \geq \inf_{(h, \ell) \in \mathcal{R}_3^2} \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0^c \right], \right. \\ & \quad \left. \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0 \right] \right\}, \end{aligned}$$

where we recall  $\mathcal{E}_0^c = \{n_2 > m - \frac{1-p}{p}s_1\}$ .

To provide a lower bound for  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0^c \right]$ , we note that under event  $\mathcal{E}_0^c$ , we have  $n_2 > m - \frac{1-p}{p}s_1$ , and hence

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0^c \right] \geq \frac{(m - \frac{1-p}{p}s_1)^+ r_2 + \min\{m, \frac{1-p}{p}s_1\} r_1}{mr_1} \geq \frac{(m - n_2)r_1}{mr_1} \geq 1 - \frac{1}{\sqrt{m}}. \quad (43)$$

Then, we give a lower bound to  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0 \right]$ . Conditional on  $\mathcal{E}_1$  and  $\mathcal{E}_0$ , we have  $\frac{1-p}{p}s_1 \geq m - n_2$  and  $\frac{1-p}{p}s_1 \leq n_1$ . That is, the number of type 1 agents in the online arrival sequence is greater than or equal to the protection level of Algorithm 1 for type 1 agents (i.e.,  $\frac{1-p}{p}s_1 \leq n_1$ ), and in addition, the algorithm ends up rejecting some of type 2 agents as  $\frac{1-p}{p}s_1 \geq m - n_2$ . Therefore, we have

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} | \mathcal{E}_2(h, \ell), \mathcal{E}_1, \mathcal{E}_0 \right] \geq \frac{(m - \frac{1-p}{p}s_1)^+ r_2 + \min\{m, \frac{1-p}{p}s_1\} r_1}{mr_1} \geq \frac{(m - n_2)r_1}{mr_1} \geq 1 - \frac{1}{\sqrt{m}},$$

where the last inequality holds because  $n_2 \leq \ell \leq \sqrt{m}$  for any  $(h, \ell) \in \mathcal{R}_3^2$ .

*Part 2: bounding  $\inf_{(h, \ell) \in \mathcal{R}_3^2} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_2(h, \ell), \mathcal{E}_1^C \right]$ .* Recall that for a fixed realization of  $\mathbf{s}$  (or equivalently  $\mathbf{n}$ ),

$$\begin{aligned} \text{REW}_A(\mathbf{n}, \rho; \mathcal{G}) &= \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\} \cdot r_2 \\ &\quad + \min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\}, n_1 \right\} \cdot r_1. \end{aligned}$$

Conditional on  $\mathcal{E}_1^C = \{n_1 \leq \frac{1-p}{p} s_1\}$ , we do not receive enough number of type 1 agents and thus, we have

$$\min \left\{ m - \min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\}, n_1 \right\} \geq \min \left\{ \max \left\{ m - n_2, \frac{1-p}{p} s_1 \right\}, n_1 \right\} = n_1.$$

That is, Algorithm 1 accepts all type 1 agents in the online arrival sequence. Then, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \rho; \mathcal{G})}{\text{OPT}(\mathbf{n})} \mid \mathcal{E}_2(h, \ell), \mathcal{E}_1^C \right] &\geq \mathbb{E}_{\mathbf{n}} \left[ \frac{\min \left\{ n_2, \left( m - \frac{1-p}{p} s_1 \right)^+ \right\} \cdot r_2 + n_1 r_1}{m r_1} \mid \mathcal{E}_2(h, \ell), \mathcal{E}_1^C \right] \\ &\geq \frac{1}{m} \mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_2(h, \ell), \mathcal{E}_1^C]. \end{aligned}$$

We will show that  $\mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_2(h, \ell), \mathcal{E}_1^C] \geq h_1$ , where  $h_1 = h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}}$  and  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ .

Since  $\mathcal{E}_1^C = \{n_1 \leq \frac{1-p}{p} s_1\} = \{h(1-p) - n_1 \geq 0\}$ , we have

$$\begin{aligned} &\mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^C, \mathcal{E}_2(h, \ell)] \\ &= h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{h(1-p) - n_1 \geq 0\}, \mathcal{E}_2(h, \ell)] \\ &= h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{h(1-p) - n_1 \geq 0\} \cap \{h(1-p) - \sqrt{h} \leq n_1 \leq h(1-p) + \sqrt{h}\}] \\ &= h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{0 \leq h(1-p) - n_1 \leq \sqrt{h}\}] \\ &\geq h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{h(1-p) - n_1 \geq 0\}] \\ &= \mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^C]. \end{aligned}$$

The inequality is because

$$\begin{aligned} &\mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{0 \leq h(1-p) - n_1 \leq \sqrt{h}\}] \\ &\leq \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{0 \leq h(1-p) - n_1 \leq \sqrt{h}\}] + \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{h(1-p) - n_1 \geq \sqrt{h}\}] \\ &= \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid \{h(1-p) - n_1 \geq 0\}]. \end{aligned}$$

Next, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} [n_1 \mid \mathcal{E}_1^C] &= h(1-p) - \mathbb{E}_{\mathbf{n}} [h(1-p) - n_1 \mid h(1-p) - n_1 \geq 0] \\ &\geq h(1-p) - \mathbb{E}_{\mathbf{s}} [|h(1-p) - n_1|] - \frac{\beta}{\sqrt{h}} \\ &\geq h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}} \\ &\geq h_0(1-p) - \sqrt{p(1-p)h_0} - \frac{\beta}{\sqrt{h_0}} \\ &= h_1, \end{aligned}$$



where the first inequality holds because by Nagaev and Chebotarev (2011), we have  $\left| \mathbb{E}_{\mathbf{n}} \left[ (h(1-p) - n_1) \mathbb{1}\{h(1-p) - n_1 \geq 0\} \right] - \mathbb{E}_{\mathbf{n}} [|h(1-p) - n_1|] \right| \leq \frac{\beta}{\sqrt{h}}$ . The second inequality follows from Berend and Kontorovich (2013) that shows  $\mathbb{E}_{\mathbf{n}} [|h(1-p) - n_1|] \leq \sqrt{p(1-p)h}$ . The third inequality holds because when  $h > h_0$ ,  $h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}}$  is a non-decreasing function for any  $p$ . To see why note that

$$\frac{\partial(h(1-p) - \sqrt{p(1-p)h} - \frac{\beta}{\sqrt{h}})}{\partial h} \geq (1-p) - \frac{1}{2} \sqrt{p(1-p)} \frac{1}{\sqrt{h}} \geq 0, \quad (44)$$

where the last inequality holds for any  $h \geq \frac{1-p}{4}$ , and we have  $h_0 > \frac{m-\sqrt{m}}{1-p} > \frac{1-p}{4}$  for all  $m \geq 2$ .

### B.5. Proof of Lemma 11

As  $\ell_0 = \min\{y : (1-p)y - \sqrt{y} \geq m\}$ , by Hoeffding's inequality, we have

$$\Pr(n_2 \geq m) = \Pr(\text{Bin}(\ell_0, 1-p) \geq m) \geq 1 - \frac{1}{\ell_0^2}.$$

Given that  $n_2 \geq m$ , the worst case is where the decision-maker accepts  $m$  type 2 agents, and the CR is lower bounded by  $\frac{mr_2}{mr_1} = \alpha$ .

#### B.5.1. Proof of Lemma 12

We have

$$\inf_{(h,\ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \right] = \inf_{(h,\ell) \in \mathcal{R}_3} \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mathbb{1}\{n_2 \geq m\} \right], \right. \\ \left. \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mathbb{1}\{n_2 < m\} \right] \right\} \quad (45)$$

We have discussed in the proof of Lemma 11 that  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mathbb{1}\{n_2 \geq m\} \right] \geq \alpha$ . Then, given Equation (45), we provide a lower bound to  $\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mathbb{1}\{n_2 < m\} \right]$ .

Conditional on event  $\{n_2 < m\}$ , the algorithm accepts all  $n_2$  type 2 agents. If  $n_2$  is larger than the protection level  $\min\{m, \frac{1-p}{p}s_2\}$ , the algorithm accepts as many type 1 agents as possible with the remaining  $m - n_2$  units of the resource. If  $n_2$  is small than the protection level  $\min\{m, \frac{1-p}{p}s_2\}$ , the algorithm rejects some type 1 agents because we can accept at most  $(m - \frac{1-p}{p}s_2)^+$  type 1 agents. Then, we have

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^C)}{\text{OPT}(\mathbf{n})} \mathbb{1}\{n_2 < m\} \right] \geq \mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \right].$$

We define the event  $\mathcal{E}_4 = \{n_2 \geq \frac{1-p}{p}s_2\}$ . Then, we have

$$\mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \right] \\ \geq \min \left\{ \mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \mathbb{1}\{\mathcal{E}_4\} \right], \right. \\ \left. \mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \mathbb{1}\{\mathcal{E}_4^C\} \right] \right\}. \quad (46)$$

Conditioned on  $\mathcal{E}_4$ , for any realization  $\mathbf{s}$ , we have

$$\frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} = \frac{n_2 r_2 + \min\{n_1, (m - n_2)\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \\ \geq \frac{mr_2}{mr_1} = \alpha.$$

When  $\mathcal{E}_4$  happens, the number of arriving type 2 agents is larger than the protection level, and the algorithm accepts all the arriving agents until there is no resource left. If OPT does not use all resources, then the algorithm accepts everyone, and hence the above ratio will be one. If OPT uses all resources, then the algorithm also uses all resources, and thus, this bound clearly holds.

Conditioned on  $\mathcal{E}_4^c$ , we have  $n_2 < \frac{1-p}{p}s_2$ . Then,

$$\begin{aligned} \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} &= \frac{n_2 r_2 + \min\{n_1, (m - \frac{1-p}{p}s_2)^+\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \\ &\geq \frac{n_2 r_2 + (m - \frac{1-p}{p}s_2)^+ r_1}{m r_1}. \end{aligned}$$

The inequality is because if  $n_1 < m - \frac{1-p}{p}s_2$ , then the algorithm accepts  $n_2$  type 2 agents and  $n_1$  type 1 agents, and the ratio is 1. Then, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + \min\{n_1, m - \max\{n_2, \min\{m, \frac{1-p}{p}s_2\}\}\} r_1}{\min\{m, n_1\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \middle| \mathcal{E}_4^c \right] &\geq \mathbb{E}_{\mathbf{n}} \left[ \frac{n_2 r_2 + (m - \frac{1-p}{p}s_2)^+ r_1}{m r_1} \middle| \mathcal{E}_4^c \right] \\ &= \mathbb{E}_{\mathbf{n}} \left[ \frac{(m - \frac{1-p}{p}s_2)^+ r_1}{m r_1} \middle| \mathcal{E}_4^c \right] \\ &\geq \mathbb{E}_{\mathbf{n}} \left[ \frac{(m - \frac{1-p}{p}s_2) r_1}{m r_1} \middle| \mathcal{E}_4^c \right] \\ &= 1 - \frac{1-p}{pm} \mathbb{E}_{\mathbf{n}} [s_2 | \mathcal{E}_4^c]. \end{aligned} \quad (47)$$

Since  $\mathcal{E}_4^c = \{s_2 \geq \ell p\}$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} [s_2 | s_2 \geq \ell p] &= \mathbb{E}_{\mathbf{n}} [s_2 - \ell p | s_2 \geq \ell p] + \ell p \\ &\leq \mathbb{E}_{\mathbf{n}} [|s_2 - \ell p|] + \frac{\beta}{\sqrt{\ell}} + \ell p \\ &\leq \sqrt{p(1-p)\ell} + \frac{\beta}{\sqrt{\ell}} + \ell p \\ &\leq \sqrt{p(1-p)\ell_0} + \frac{\beta}{\sqrt{m}} + \ell_0 p \\ &= \ell_1, \end{aligned}$$

where the first inequality holds because by Nagaev and Chebotarev (2011), we have  $|\mathbb{E}_{\mathbf{n}} [s_2 - \ell p | s_2 \geq \ell p] - \mathbb{E}_{\mathbf{n}} [|s_2 - \ell p|]| \leq \frac{\beta}{\sqrt{\ell}}$ . The second inequality follows from Berend and Kontorovich (2013) that shows  $\mathbb{E}_{\mathbf{n}} [|s_2 - \ell p|] \leq \sqrt{p(1-p)\ell}$ .

Finally, we have

$$\inf_{(h, \ell) \in \mathcal{R}_3} \mathbb{E}_{\mathbf{n}} \left[ \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G}^c)}{\text{OPT}(\mathbf{n})} \right] \geq \min \left\{ \alpha, 1 - \frac{1-p}{pm} \ell_1 \right\}.$$

## Appendix C: Proof of Proposition 2

To show the asymptotic result, we only need to consider the case that  $m \geq m_1$ . We can write the CR as

$$\text{CR}_A \geq \min \{ \text{CR}_1, \text{CR}_2, \overline{\text{CR}}_3 \}.$$

To show the result, we will show how  $\text{CR}_1, \text{CR}_2$ , and  $\overline{\text{CR}}_3$  scale with  $m$ . First it is easy to see that  $\text{CR}_1 = \min \left\{ \left(1 - \frac{1-p}{p\sqrt{m}}\right)^+, 1 - \frac{r_1 - r_2}{(r_1 - r_2) + r_2 \sqrt{m}} \right\}$  scales with  $1 - \Theta(1/(p\sqrt{m}))$ . Similarly, it is easy to see that  $\text{CR}_2 =$

$\min \left\{ \left(1 - \frac{1}{\sqrt{m}}\right) \left(1 - \frac{1}{m}\right), \frac{h_1}{m} \right\}$  scales with  $1 - \Theta(1/\sqrt{m})$ . This is because  $h_0 = \frac{m - \sqrt{m}}{1 - p} + \frac{1}{2p^2} + \sqrt{\frac{1}{4p^4} + \frac{(m - \sqrt{m})}{p^2(1 - p)}}$ , and  $h_1 = h_0(1 - p) - \sqrt{p(1 - p)h_0} - \frac{\beta}{\sqrt{h_0}} = m - (1 + \sqrt{p})\sqrt{m} - o(\sqrt{m})$ . Then, we have  $\frac{h_1}{m} = 1 - \Theta(1/\sqrt{m})$ .

To complete the proof, we show how

$$\overline{\text{CR}}_3 = \left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2 W$$

scales with  $m$ . Observe that  $\left(1 - \frac{1}{m}\right)^2 \left(1 - \frac{1}{(\sqrt{mp} - m^{1/4})^2}\right)^2 = 1 - \Theta(1/(p^2m))$ . When  $p = \omega(1/\sqrt{m})$ ,  $\Theta(1/(p^2m))$  is dominated by  $\Theta(1/(p\sqrt{m}))$ . Further, by the analysis for  $\text{CR}_2$ , we know that  $W = \min \left\{ 1 - \frac{1}{\sqrt{m}}, \frac{h_1}{m} \right\}$  also scales with  $1 - \Theta(1/\sqrt{m})$ . Putting these together, we conclude that  $\text{CR}_A = 1 - \Theta(1/(p\sqrt{m}))$ , as desired.

## Appendix D: Proof of Theorem 2

We construct the following input family  $\mathbf{F}$ : let  $\underline{h} = 0$  and  $\bar{h} = pm$ . The input family  $\mathbf{F}$  contains all  $(h, \ell)$  such that  $h \in [\underline{h}, \bar{h}]$  and  $\ell = \frac{10000 \cdot m}{p}$ . For any  $h \in [\underline{h}, \bar{h}]$ , we then denote  $I_h$  as a random arrival sequence under which  $n_2$  type 2 agents arrive followed by  $n_1$  type 1 agents, where we recall that  $n_1 \sim \text{Bin}(h, 1 - p)$ ,  $n_2 \sim \text{Bin}(\ell, 1 - p)$ . We denote  $I_h$  as the random arrival sequences, where in  $I_h$ ,  $n_2$  type 2 agents arrive followed by  $n_1$  type 1 agents.

Our goal is to characterize an upper bound on the CR of any deterministic algorithm under the family  $\mathbf{F}$ . Consider a specific input  $h \in [\underline{h}, \bar{h}]$ . The decision maker gets the sample  $(s_1, s_2)$  at the beginning of the allocation period, where  $s_1 \sim \text{Bin}(h, p)$ , and  $s_2 \sim \text{Bin}(\ell, p)$ . As we select  $\ell$  to be a super large constant, we have with probability 1,  $s_2 \gg m$ . Therefore, under the family  $\mathbf{F}$ , the decision-maker knows that there will be more than  $m$  type 2 agents showing up. Based on the structure that all type 2 agents arrive before all type 1 agents, any deterministic algorithm has to decide how many type 2 agents they accept if they observe  $(s_1, s_2)$ . As  $s_2$  is always larger than  $m$ , it does not impact the decisions. Put differently, any deterministic algorithm can be represented by a mapping  $z(\cdot): [0, \bar{h}] \rightarrow [0, m]$ , such that once they observe  $(s_1, s_2)$ , they will accept  $m - z(s_1)$  type 2 agents. (As the number of type 1 agents,  $h$ , is less than  $m$ , the number of type 1 agents in the online arrival sequence  $n_1 < m$ .) Let  $A$  be an algorithm associated with mapping  $z$ . We define the following loss function for algorithm  $A$  under arrival sequence  $I$  and sample information  $\mathbf{s}$ :

$$\mathbf{L}_A(I, \mathbf{s}) = \text{OPT}(I) - \text{REW}_A(I, \mathbf{s}). \quad (48)$$

(Note that since  $r_1$  and  $r_2$  are known, we can replace  $\psi = (\mathbf{s}, \boldsymbol{\rho})$  with  $\mathbf{s}$ .) In the rest of this proof, we use the shorthand of  $\mathbf{L}(I, \mathbf{s}, z)$  and  $\text{REW}(I, \mathbf{s}, z)$  in place of  $\mathbf{L}_A(I, \mathbf{s})$  and  $\text{REW}_A(I, \mathbf{s})$ , respectively. Then, the CR of a deterministic algorithm  $A$  with mapping  $z(\cdot)$  on family  $\mathbf{F}$  is at least

$$\text{CR}(z) = 1 - \sup_{(h, \ell) \in \mathbf{F}} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbf{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right]. \quad (49)$$

In what follows, we provide an upper bound on  $\text{CR}(z)$  in Equation (49). To do so, we consider the following two cases:

- *Case 1:*  $n_1 \geq z(s_1)$ . In this case, the number of type 1 agent arriving online (which is a random variable) is larger than the remaining number of resources after we accept  $m - z(s_1)$  type 2 agents. Therefore, the decision-maker accepts  $m - z(s_1)$  type 2 agents and  $z(s_1)$  type 1 agents. The loss (i.e., the difference between the optimal clairvoyant cumulative expected reward and what the decision-maker gets which is defined in Equation (48)) is then  $(n_1 - z(s_1))(r_1 - r_2)$ .

- *Case 2:*  $n_1 < z(s_1)$ . In this case, the number of type 1 agent arriving online is less than the remaining number of resources after we accept  $m - z(s_1)$  type 2 agents. Therefore, the decision-maker accepts  $m - z(s_1)$  type 2 agents and  $n_1$  type 1 agents. Therefore, we waste  $z(s_1) - n_1$  units of resources. In this case, the loss, defined in Equation (48), is  $(z(s_1) - n_1)r_2$ .

Putting these two cases together, we have

$$\begin{aligned} \mathbf{L}(I, \mathbf{s}, z) &= (n_1 - z(s_1))(r_1 - r_2)\mathbf{1}(n_1 \geq z(s_1)) + (z(s_1) - n_1)r_2\mathbf{1}(n_1 < z(s_1)) \\ &= (n_1 - z(s_1))r_1\mathbf{1}(n_1 \geq z(s_1)) + (z(s_1) - n_1)r_2. \end{aligned}$$

Hence, an upper bound on the CR of any deterministic algorithm on family  $\mathbf{F}$ , denoted by  $\overline{\text{CR}}$ , is at least

$$\begin{aligned} \overline{\text{CR}} &= \max_z \text{CR}(z) \\ &= 1 - \min_z \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbf{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] \\ &= 1 - \min_z \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{(n_1 - z(s_1))r_1\mathbf{1}(n_1 \geq z(s_1)) + (z(s_1) - n_1)r_2}{n_1r_1 + (m - n_1)r_2} \right], \end{aligned} \quad (50)$$

where  $\text{CR}(z)$  is defined in Equation (49).

Although we can solve this minmax optimization problem by some solver because the objective function is convex in every dimension, we cannot calculate the value of the optimal solution in a closed form. Therefore, we develop the analysis below to show that  $\overline{\text{CR}}$  is  $1 - \Theta(\frac{1}{p\sqrt{m}})$ .

In our analysis, we first focus on a specific mapping:  $z(s_1) = \frac{1-p}{p}s_1$  for any  $s_1 \in [0, \bar{h}]$ . (Since  $s_1 \leq \bar{h} = mp$ , we have  $z(s_1) \leq m$ .) For this mapping, we calculate the CR of its associated deterministic algorithm. Second, we compare any other feasible mappings with this mapping to show our result. We want to highlight that this mapping (i.e.,  $z(s_1) = \frac{1-p}{p}s_1$ ) is not the optimal mapping: The optimal mapping should depend on  $r_1$  and  $r_2$ , but we cannot find its closed form. This mapping is considered as a benchmark with which we compare all other mappings.

**LEMMA 13 (The CR under Mapping  $z(\cdot)$ ).** *Consider the following mapping  $z : [0, \bar{h}] \rightarrow [0, m]$  such that  $z(s_1) = \frac{1-p}{p}s_1$  for any  $s_1 \in [0, \bar{h}]$ . Then, for any value of  $m$*

$$\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{L(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] \leq \frac{r_1\sqrt{1-p}}{r_2} \frac{1}{p\sqrt{m}} + \frac{\sqrt{2}\beta r_1}{p\sqrt{p}r_2} \frac{1}{m\sqrt{m}},$$

and

$$\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{L(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] \geq \frac{1}{2\sqrt{2}} \sqrt{1-p} \frac{1}{p\sqrt{m}} - \frac{\beta}{p\sqrt{p}} \frac{1}{m\sqrt{m}},$$

where  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ . This implies that  $\text{CR}(z) = 1 - \Theta(\frac{1}{p\sqrt{m}})$ .

By Lemma 13, we have under that particular mapping  $z$ , we have  $\text{CR}(z) = 1 - \Theta(\frac{1}{p\sqrt{m}})$  over input family  $\mathbf{F}$ . The next lemma shows that there does not exist any other mapping  $\hat{z}(\cdot)$  such that the CR over  $\mathbf{F}$  is  $1 - o(\frac{1}{p\sqrt{m}})$ .

**LEMMA 14 (Other Mappings).** *Consider the setting in Section D and the input family  $\mathbf{F}$ . For this setting, there does not exist any mapping  $\hat{z}(\cdot)$  such that the competitive ratio over  $\mathbf{F}$  is  $1 - o(\frac{1}{p\sqrt{m}})$ .*

To show this, we compare all other mappings  $\hat{z}: [0, \bar{h}] \rightarrow [0, m]$ , with  $z(s_1) = \frac{1-p}{p}s_1$ , and using contradiction, we assume that there exists a mapping  $\hat{z}$  which achieves the CR of  $1 - o(\frac{1}{p\sqrt{m}})$ . As  $z$  achieves a CR of  $1 - \Theta(\frac{1}{p\sqrt{m}})$ , the difference between the CR under  $\hat{z}$  and  $z$  is at least  $\Theta(\frac{1}{p\sqrt{m}})$ . This gives constraints that  $\hat{z}$  and  $z$  should be different on a set, and the difference between  $\hat{z}$  and  $z$  on the points in this set is lower bounded. Then, we further claim that for a mapping  $\hat{z}$  which has a certain difference compared to  $z$ , the CR is at least  $1 - \Omega(\frac{1}{p\sqrt{m}})$ , which gives a contradiction and shows the best algorithm (mapping) can only achieve an asymptotic CR of  $1 - \Theta(\frac{1}{p\sqrt{m}})$ .

### D.1. Proof of Lemma 13

The proof has two parts. In the first part, we present an upper bound on  $\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbf{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right]$  and in the second part, we present a lower bound on the same quantity.

*First part: upper bound.* By definition,  $z(s_1) = \frac{1-p}{p}s_1$  for any  $s_1 \in [0, \bar{h}]$ , and hence we have  $n_1 - z(s_1) = h - s_1 - \frac{1-p}{p}s_1 = h - \frac{1}{p}s_1$ . Similarly,  $z(s_1) - n_1 = \frac{1}{p}s_1 - h$ . Then, by Equation (50), we have

$$\begin{aligned} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbf{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] &\leq \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{(n_1 - z(s_1))r_1 \mathbf{1}(n_1 \geq z(s_1)) + (z(s_1) - n_1)r_2}{mr_2} \right] \\ &= \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \left( h - \frac{1}{p}s_1 \right) \frac{r_1}{r_2} \mathbf{1}\left(h - \frac{1}{p}s_1 \geq 0\right) + \left( \frac{1}{p}s_1 - h \right) \right] \\ &= \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \left( h - \frac{1}{p}s_1 \right) \frac{r_1}{r_2} \mathbf{1}\left(h - \frac{1}{p}s_1 \geq 0\right) \right] \\ &= \frac{r_1}{mpr_2} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} [(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)]. \end{aligned}$$

The second to last equation follows because  $\mathbb{E}_{\mathbf{s}}[\frac{1}{p}s_1 - h] = 0$ . Recall that  $s_1 \sim \text{Bin}(h, p)$ . Next, we bound  $\mathbb{E}_{\mathbf{s}}[(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)]$ . We will show that

$$\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} [(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)] \leq \frac{1}{2} \sqrt{1-p} \sqrt{m} + \frac{\sqrt{2}\beta}{\sqrt{pm}}, \quad (51)$$

where  $\beta = 0.4215 \cdot \frac{p^2 + (1-p)^2}{p(1-p)}$ . This confirms that

$$\begin{aligned} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_{\mathbf{s}} \left[ \frac{\mathbf{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] &\leq \frac{r_1}{mpr_2} \left( \frac{1}{2} \sqrt{1-p} \sqrt{m} + \frac{\sqrt{2}\beta}{\sqrt{pm}} \right) \\ &= \frac{r_1 \sqrt{1-p}}{r_2 p \sqrt{m}} + \frac{\sqrt{2}\beta r_1}{p \sqrt{p} r_2} \frac{1}{m \sqrt{m}}, \end{aligned}$$

which is the desired result.

It remains to show Equation (51). By Nagaev and Chebotarev (2011), we have

$$\left| \mathbb{E}_{\mathbf{s}} [(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)] - \frac{1}{2} \mathbb{E}_{\mathbf{s}} [|hp - s_1|] \right| \leq \frac{\beta}{\sqrt{h}}. \quad (52)$$

In addition, Berend and Kontorovich (2013) provides a sharp estimate of  $\mathbb{E}_{\mathbf{s}} [|hp - s_1|]$ :

$$\frac{1}{\sqrt{2}} \sqrt{\frac{1-p}{p}} h \leq \mathbb{E}_{\mathbf{s}} [|hp - s_1|] \leq \sqrt{\frac{1-p}{p}} h. \quad (53)$$

This leads to

$$\mathbb{E}_{\mathbf{s}} [(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)] \leq \frac{1}{2} \mathbb{E}_{\mathbf{s}} [|hp - s_1|] + \frac{\beta}{\sqrt{h}} \leq \frac{1}{2} \sqrt{\frac{1-p}{p}} h + \frac{\beta}{\sqrt{h}}. \quad (54)$$

Therefore, for large enough  $m$ , we have

$$\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s [(hp - s_1) \mathbf{1}(hp - s_1 \geq 0)] \leq \sqrt{\frac{1-p}{p}} \bar{h} + 0.4215 \frac{\beta}{\sqrt{\bar{h}}} = \frac{1}{2} \sqrt{1-p} \sqrt{m} + \frac{\sqrt{2}\beta}{\sqrt{pm}}, \quad (55)$$

which is the desired result.

*Second part: lower bound.* Here, we present a lower bound on  $\max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \frac{\mathbb{L}(I, \mathbf{s}, \hat{z})}{\text{OPT}(I)} \right]$ , following similar steps in the first part. By definition,

$$\begin{aligned} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \frac{\mathbb{L}(I, \mathbf{s}, z)}{\text{OPT}(I)} \right] &\geq \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \frac{(n_1 - z(s_1))r_1 \mathbf{1}(n_1 \geq z(s_1)) + (z(s_1) - n_1)r_2}{mr_1} \right] \\ &\geq \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \left( h - \frac{1}{p}s_1 \right) \mathbf{1}\left( h - \frac{1}{p}s_1 \geq 0 \right) + \left( \frac{1}{p}s_1 - h \right) \right] \\ &= \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \left( h - \frac{1}{p}s_1 \right) \frac{r_1}{r_2} \mathbf{1}\left( h - \frac{1}{p}s_1 \geq 0 \right) \right] \\ &= \frac{1}{mp} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s [(hp - s_1) \mathbf{1}(hp - s_1 \leq 0)] \\ &\geq \frac{1}{mp} \max_{h \in [\underline{h}, \bar{h}]} \left( \frac{1}{2} \mathbb{E}_s [ |hp - s_1| ] - \frac{\beta}{\sqrt{h}} \right) \\ &\geq \frac{1}{mp} \max_{h \in [\underline{h}, \bar{h}]} \left( \frac{1}{2\sqrt{2}} \sqrt{\frac{1-p}{p}} h - \frac{\beta}{\sqrt{h}} \right) \\ &= \frac{1}{mp} \left( \frac{1}{2\sqrt{2}} \sqrt{(1-p)m} - \frac{\beta}{\sqrt{pm}} \right) \\ &= \frac{1}{2\sqrt{2}} \sqrt{(1-p)} \frac{1}{p\sqrt{m}} - \frac{\beta}{p\sqrt{p}} \frac{1}{m\sqrt{m}}. \end{aligned}$$

The second to last inequality is due to Equation (52). The last inequality is due to Equation (53).

## D.2. Proof of Lemma 14

Consider any arbitrary mapping  $\hat{z}(\cdot)$ . By Equation (50), we have

$$\begin{aligned} \text{CR}(\hat{z}) &= 1 - \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \frac{\mathbb{L}(I, \mathbf{s}, \hat{z})}{\text{OPT}(I)} \right] \\ &\leq 1 - \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ \frac{(n_1 - \hat{z}(s_1))r_1 \mathbf{1}(n_1 \geq \hat{z}(s_1)) + (\hat{z}(s_1) - n_1)r_2}{mr_2} \right] \\ &= 1 - \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ (n_1 - \hat{z}(s_1)) \frac{r_1}{r_2} \mathbf{1}(n_1 \geq \hat{z}(s_1)) + (\hat{z}(s_1) - n_1) \right]. \end{aligned}$$

Let  $\hat{y}(s_1) = s_1 + \hat{z}(s_1)$  for any  $s_1 \in [0, \bar{h}]$ . Then, since  $n_1 + s_1 = h$ , we have

$$\text{CR}(\hat{z}) \leq 1 - \frac{1}{m} \max_{h \in [\underline{h}, \bar{h}]} \mathbb{E}_s \left[ (h - \hat{y}(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq \hat{y}(s_1)) + (\hat{y}(s_1) - h) \right].$$

Our goal is to show that  $\text{CR}(\hat{z}) = 1 - \Omega\left(\frac{1}{p\sqrt{m}}\right)$ . We show this by contradiction. Contrary to our claim, suppose that there exists a feasible mapping  $y_0(\cdot)$  with  $y_0(s_1) = s_1 + z_0(s_1)$  under which

$$\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_s \left[ (h - y_0(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq y_0(s_1)) + (y_0(s_1) - h) \right]}{p\sqrt{m}} = 0. \quad (56)$$

Note that if Equation (56) holds, we have  $\text{CR}(z_0) = 1 - o\left(\frac{1}{p\sqrt{m}}\right)$ . Now observe that  $((h - y_0(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq y_0(s_1))) \geq 0$ , and hence if Equation (56) holds, we must have

$$\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_s [y_0(s_1) - h]}{p\sqrt{m}} \leq 0. \quad (57)$$

Next we consider two cases. In the first case, we assume that  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{\mathbf{s}}[y_0(s_1) - h]}{p\sqrt{m}} < 0$  and we reach a contradiction under this assumption. In the second case, assume that  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{\mathbf{s}}[y_0(s_1) - h]}{p\sqrt{m}} = 0$  and again reach a contradiction.

*Case 1:*  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{\mathbf{s}}[y_0(s_1) - h]}{p\sqrt{m}} < 0$ . Suppose that  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{\mathbf{s}}[y_0(s_1) - h]}{p\sqrt{m}} = \eta < 0$ . Then, we show that  $y_0(\cdot)$  does not satisfy Equation (56), which is a contradiction. The left hand side of Equation (56) can be written as

$$\begin{aligned} & \lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h, p)} \left[ (h - y_0(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq y_0(s_1)) + (y_0(s_1) - h) \right]}{p\sqrt{m}} \\ & \geq \lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h, p)} \left[ (h - y_0(s_1)) \frac{r_1}{r_2} + (y_0(s_1) - h) \right]}{p\sqrt{m}} = -\frac{r_1}{r_2} \eta + \eta = -\frac{r_1 - r_2}{r_2} \eta > 0, \end{aligned}$$

which implies that Equation (56) does not hold and hence it is a contradiction.

*Case 2:*  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_{\mathbf{s}}[y_0(s_1) - h]}{p\sqrt{m}} = 0$ . Here, we compare  $y_0(\cdot)$  and  $y(\cdot)$  where we recall  $y(s_1) = s_1 + z(s_1) = \frac{1}{p} s_1$  for any  $s_1 \in [0, \bar{h}]$ . We then show a contradiction. In Lemma 13, we have shown that if  $y(s_1) = \frac{1}{p} s_1$  (or equivalently  $z(s_1) = \frac{1-p}{p} s_1$ ), then for any  $h \in [\underline{h}, \bar{h}]$ , we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h, p)} \left[ (h - y(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq y(s_1)) + (y(s_1) - h) \right]}{p\sqrt{m}} \\ & = \lim_{m \rightarrow \infty} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h, p)} \left[ \left( h - \frac{s_1}{p} \right) \frac{r_1}{r_2} \mathbf{1}\left( h \geq \frac{s_1}{p} \right) + \left( \frac{s_1}{p} - h \right) \right]}{p\sqrt{m}} > 0. \end{aligned} \quad (58)$$

Let Equation (56) takes its maximum at  $h = h_0 \in [\underline{h}, \bar{h}]$ , and notice that  $h_0$  is a function with respect to  $m$ . Also, note that Equation (58) holds for any  $h$ , including  $h_0$ . Hence, by subtracting Equation (58) from Equation (56) evaluated at  $h_0$ , we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} \left[ (h_0 p - s_1) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq s_1) + (s_1 - h_0 p) \right] \right. \\ & \quad \left. - \mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} \left[ (h_0 p - y_0(s_1) p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(s_1)) + (y_0(s_1) - h_0) \right] \right) > 0. \end{aligned} \quad (59)$$

Now recall that under this case, we have  $\lim_{m \rightarrow \infty} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} [y_0(s_1) p - h_0 p]}{p\sqrt{m}} = 0$  where  $s_1 \sim \text{Bin}(h_0, p)$ . This implies that  $\lim_{m \rightarrow \infty} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} [s_1 - y_0(s_1) p]}{\sqrt{m} p} = 0$ , as  $\mathbb{E}[s_1] = h_0 p$ . Therefore, Equation (59) can be written as

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} \left[ (h_0 p - s_1) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq s_1) - (h_0 p - y_0(s_1) p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(s_1)) \right] \right) > 0. \quad (60)$$

Next, let sequence  $q_d(m) = \frac{|y_0(d) p - d|}{p\sqrt{m}}$  for  $d \in [0, \bar{h}]$ , We define set  $\mathcal{D}$  as follows

$$\mathcal{D} = \{d \in [0, \bar{h}] : \lim_{m \rightarrow \infty} q_d(m) > 0\}.$$

By definition  $q_d(m)$  is non-negative for all  $d$ , and hence we define the complement of set  $\mathcal{D}$  as follows

$$\mathcal{D}^c = \{d \in [0, \bar{h}] : \lim_{m \rightarrow \infty} q_d = 0\}. \quad (61)$$

Then, we can write the inner term of Equation (60) as

$$\begin{aligned} \mathbb{E}_{s_1 \sim \text{Bin}(h_0, p)} & \left[ (h_0 p - s_1) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq s_1) - (h_0 p - y_0(s_1)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(s_1)) \right] \\ & = \sum_{d \in \mathcal{D}} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \right] \\ & \quad + \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \right]. \end{aligned}$$

In Step (i), we show that

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \right] = 0. \quad (62)$$

In Step (ii), we show that

$$\lim_{m \rightarrow \infty} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) > 0.$$

By these two equations, we can make a contradiction of Equation (57).

*Step (i):* Observe that  $\sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) \leq 1$ , and hence we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)p) \right] \\ & \leq \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{r_1}{r_2} \max_{d \in \mathcal{D}^c} [(h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 \geq y_0(d))]. \end{aligned} \quad (63)$$

The inner term of the above equation can then be written as

$$\begin{aligned} & (h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 p \geq y_0(d)p) \\ & = ((h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}_{h_0 p \geq y_0(d)p}) \mathbf{1}(d < y_0(d)p) \\ & \quad + ((h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 p \geq y_0(d)p)) \mathbf{1}(d \geq y_0(d)p). \end{aligned}$$

Notice that

$$\begin{aligned} & ((h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 \geq y_0(d))) \mathbf{1}(d < y_0(d)p) \\ & = ((h_0 p - d - h_0 p + y_0(d)p) \mathbf{1}(h_0 p \geq y_0(d)p) + (h_0 p - d) \mathbf{1}(d \leq h_0 p \leq y_0(d)p)) \mathbf{1}(d < y_0(d)p) \\ & = ((y_0(d)p - d) \mathbf{1}(h_0 p \geq y_0(d)p) + (h_0 p - d) \mathbf{1}(d \leq h_0 p \leq y_0(d)p)) \mathbf{1}(d < y_0(d)p) \\ & \leq (|y_0(d)p - d| + (y_0(d)p - d)) \mathbf{1}(d < y_0(d)p) \\ & \leq 2(|y_0(d)p - d|) \mathbf{1}(d < y_0(d)p). \end{aligned}$$

Similarly, we have

$$((h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 \geq y_0(d))) \mathbf{1}(d \geq y_0(d)p) \leq 2(|y_0(d)p - d|) \mathbf{1}(d \geq y_0(d)p).$$

Therefore, the inner term of Equation (63) is upper bounded by

$$(h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 p \geq y_0(d)p) \leq 2|y_0(d)p - d|. \quad (64)$$



Using this in Equation (64) gives us

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{r_1}{r_2} \max_{d \in \mathcal{D}^c} [(h_0 p - d) \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \mathbf{1}(h_0 \geq y_0(d))] \\ & \leq \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{r_1}{r_2} \max_{d \in \mathcal{D}^c} [2|y_0(d)p - d|] \\ & = \lim_{m \rightarrow \infty} \frac{2r_1}{r_2} \max_{d \in \mathcal{D}^c} \frac{|y_0(d)p - d|}{p\sqrt{m}}, \\ & = 0, \end{aligned}$$

which is what we wanted to show. Note that the last equation holds because of definition of  $\mathcal{D}^c$  in Equation (61).

*Step (ii):* By Equation (60) and the inequality above we have

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \right] > 0. \quad (65)$$

Notice that for  $d \in \mathcal{D}$ , we have  $\lim_{m \rightarrow \infty} \frac{|y_0(d)p - d|}{p\sqrt{m}} > 0$ . We define  $c_1, c_2 > 0$  be the constant such that

$$\sup_{d \in \mathcal{D}} \lim_{m \rightarrow \infty} \frac{|y_0(d)p - d|}{p\sqrt{m}} = c_1 \quad \text{and} \quad \inf_{d \in \mathcal{D}} \lim_{m \rightarrow \infty} \frac{|y_0(d)p - d|}{p\sqrt{m}} = c_2.$$

Recall that  $\mathcal{D} = \{d \in \{0, 1, \dots\} : \lim_{m \rightarrow \infty} q_d(m) > 0\}$ , where  $q_d(m) = \frac{|y_0(d)p - d|}{p\sqrt{m}}$  for  $d \in \{0, 1, \dots\}$ .

By Equation (64), we have

$$(h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \leq 2 \frac{r_1}{r_2} |y_0(d)p - d|.$$

Therefore, Equation (65) can be upper bounded as

$$0 < \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) 2 \frac{r_1}{r_2} |y_0(d)p - d| \leq 2c_1 \frac{r_1}{r_2} \lim_{m \rightarrow \infty} \sum_{d \in \mathcal{D}} \Pr(s_1 = d), \quad (66)$$

which implies that

$$\lim_{m \rightarrow \infty} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) > 0. \quad (67)$$

We are now ready to show the contraction. Recall that in the current case, we assumed that  $\lim_{m \rightarrow \infty} \frac{\mathbb{E}_{\mathbf{s}}(y_0(s_1) - h_0)}{p\sqrt{m}} = 0$ , which implies that  $\lim_{m \rightarrow \infty} \frac{\mathbb{E}_{\mathbf{s}}|y_0(s_1) - h_0|}{p\sqrt{m}} = 0$ . As in Steps (i) and (ii), we have shown that  $\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) \left[ (h_0 p - d) \frac{r_1}{r_2} \mathbf{1}(h_0 p \geq d) - (h_0 p - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(h_0 \geq y_0(d)) \right] = 0$  and  $\lim_{m \rightarrow \infty} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) > 0$ , by the following chain of equations we have the desired contradiction.

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{\mathbb{E}_{\mathbf{s}}|y_0(s_1) - h_0|}{p\sqrt{m}} &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \sum_{d \in \mathcal{D}} \Pr(s_1 = d) |y_0(d)p - h_0 p| + \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) |y_0(d)p - h_0 p| \right) \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \sum_{d \in \mathcal{D}} \Pr(s_1 = d) |y_0(d)p - d + d - h_0 p| \right. \\ &\quad \left. + \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) |y_0(d)p - d + d - h_0 p| \right) \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \sum_{d \in \mathcal{D}} \Pr(s_1 = d) |y_0(d)p - d + d - h_0 p| + \sum_{d \in \mathcal{D}^c} \Pr(s_1 = d) (d - h_0 p) \right) \end{aligned}$$

$$\begin{aligned}
&\geq \lim_{m \rightarrow \infty} \frac{1}{\sqrt{m}} \left( \sum_{d \in [h_0]} (d - h_0 p) + \sum_{d \in \mathcal{D}} \Pr(s_1 = d) |y_0(d)p - d| \right) \\
&= \mathbb{E}_s [s_1 - h_0 p] + \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) |y_0(d)p - d| \\
&\geq \lim_{m \rightarrow \infty} \sum_{d \in \mathcal{D}} \Pr(s_1 = d) c_2 \\
&> 0.
\end{aligned}$$

The first inequality holds because  $\mathcal{D}^c = \{d \in [0, \bar{h}] : \lim_{m \rightarrow \infty} q_d(m) = 0\}$ , where  $q_d(m) = \frac{|y_0(d)p - d|}{p\sqrt{m}}$  for  $d \in [0, \bar{h}]$ . The third equation is due to definition of  $\mathcal{D}^c$ . The fourth inequality is because  $|y_0(d)p - d + d - h_0 p| \geq |y_0(d)p - d| + (d - h_0 p)$ . The last inequality is from Equation (67). This is a contradiction to  $\lim_{m \rightarrow \infty} \max_{h \in [\underline{h}, \bar{h}]} \frac{\mathbb{E}_s [y_0(s_1) - h]}{p\sqrt{m}} = 0$ .

Put the two cases together, we conclude that there does not exist such mapping  $y_0(\cdot)$  with which the competitive ratio on family of  $\mathbf{F}$  is  $1 - o(\frac{1}{p\sqrt{m}})$ .

### Appendix E: Proof of Theorem 3

We construct the following random input distribution by taking advantage of input family  $\mathbf{F}$  that we used in the proof of Theorem 2. Recall that under a given input  $h, \ell$  in family  $\mathbf{F}$ ,  $I_h$  is the random arrival sequence such that  $n_2$  type 2 agents arrive first followed by  $n_1$  type 1 agents, where  $n_1 \sim \text{Bin}(h, 1 - p)$ ,  $n_2 \sim \text{Bin}(\ell, 1 - p)$ , and  $\ell = \frac{10000 \cdot m}{p}$ , and  $h \in [\underline{h}, \bar{h}]$  with  $\underline{h} = \frac{1}{2}pm$  and  $\bar{h} = pm$ . Then, in our random input distribution, we choose one of the feasible  $h \in [\underline{h}, \bar{h}]$  uniformly at random; that is  $\Pr(\text{Uniform}[\underline{h}, \bar{h}] = h) = \frac{1}{\bar{h} - \underline{h} + 1}$  for any  $h \in [\underline{h}, \bar{h}]$ .

Observe that due to our sampling procedure, even when the input  $h \in [\underline{h}, \bar{h}]$  is realized, the online arrival sequence  $I = I_h$  is still random. That prevents us from applying the Von Neuman/Yao principle Seiden (2000) to our setting. At least in a footnote, clarify what the Von Neuman/Yao principle is and why we cannot use it in our setting. Nonetheless, we derive a result similar to the Von Neuman/Yao principle that can be applied to our setting.

**LEMMA 15 (Von Neuman/Yao principle: from deterministic to randomized inputs).** *Consider a setting where the adversary chooses a meta input  $X \in \mathcal{X}$ , and then based on the meta input  $X$ , a random input  $\tilde{X}$  from a distribution  $D(X)$  is realized. For any random or deterministic algorithm  $A$  and meta input  $X \in \mathcal{X}$ , let  $C_A(\tilde{X}) \in [0, 1]$  be the (realized) reward of algorithm  $A$  under input  $\tilde{X}$  over the reward of the optimal in-hindsight algorithm that knows  $\tilde{X}$  in advance, where  $\tilde{X} \sim D(X)$ . Then,  $\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]$  is the competitive ratio of algorithm  $A$  under meta input  $X$ , where the expectation is with respect to  $\tilde{X}$  and any randomness in algorithm  $A$ .*

Now, let  $\mathcal{A}_d$  be the set of all deterministic algorithms. Let  $P$  be a probability distribution over any deterministic algorithms. Define  $A \sim P$  as an algorithm chosen according to probability distribution  $P$ .<sup>15</sup> Further, let  $Q$  be a probability distribution over the meta inputs in  $\mathcal{X}$ , and let  $X \sim Q$  denote a random meta input

<sup>15</sup> By Yao (1977), Ball and Queyranne (2009), any randomized algorithm may be viewed as a random choice  $A \sim P$  among deterministic algorithms.

chosen according to  $Q$ . Then, for any distribution  $P$  over algorithms, distribution  $Q$  over meta inputs, we have

$$\min_{\tilde{X} \in \tilde{\mathcal{X}}} \mathbb{E}_{A \sim P} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]] \leq \max_{A \in \mathcal{A}_d} \mathbb{E}_{X \sim Q} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]]. \quad (68)$$

The proof can be found in Section E.1. We now ready to show the result using Lemma 15. Here, our meta inputs are  $(h, \ell) \in \mathbf{F}$  and our inputs are online arrival sequence  $I_h$ . Furthermore, the distribution over our meta inputs, i.e.,  $Q$  in Lemma 15, is uniform distribution. Recall that  $\Pr(\text{Uniform}[\underline{h}, \bar{h}] = h) = \frac{1}{\bar{h} - \underline{h} + 1}$  for any  $h \in [\underline{h}, \bar{h}]$ .

Our goal here is to upper bound the competitive ratio of any randomized algorithm on family  $\mathbf{F}$ , defined above. That is, for any distribution  $P$  over any feasible deterministic algorithms  $A$ , we would like to bound the following quantity

$$\text{CR}_{\text{rand}} := \min_{\hat{I} \in \mathbf{F}} \mathbb{E}_{A \sim P} [\mathbb{E}_{\hat{I}} [\text{CR}_A(I)]] .$$

By Lemma 15, we have

$$\text{CR}_{\text{rand}} \leq \max_{A \in \mathcal{A}_d} \mathbb{E}_{\hat{I} \sim Q} [\mathbb{E}_{\hat{I}} [\text{CR}_A(I)]] = \max_{A \in \mathcal{A}_d} \mathbb{E}_{\hat{I} \sim Q} \left[ \mathbb{E}_{\hat{I}} \left[ \frac{\text{REW}_A(I)}{\text{OPT}(I)} \right] \right],$$

where  $\mathcal{A}_d$  is the set of all deterministic algorithms. Since  $Q$  is a uniform distribution over any  $h \in [\underline{h}, \bar{h}]$ , we have

$$\text{CR}_{\text{rand}} \leq \frac{1}{\bar{h} - \underline{h} + 1} \cdot \max_{A \in \mathcal{A}_d} \sum_{h=\underline{h}}^{\bar{h}} \mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ \frac{\text{REW}_A(I_h, s_1)}{\text{OPT}(I_h)} \right], \quad (69)$$

where  $I_h$  is an arriving instance that a huge number of type 2 agents arrive first, and followed by  $h$  type 1 agents.  $\hat{I}_h$  is a random variable which is consisted of  $I_h$  with a uniform distribution on  $h \in [\underline{h}, \bar{h}]$ . Recall that under  $\hat{I}_h$ ,  $\text{REW}_A$  does not depend on  $s_2 \gg m$ , and hence we denote  $\text{REW}_A(I, \mathbf{s})$  with  $\text{REW}_A(I, s_1)$ . We would like to show that for any deterministic algorithm  $A \in \mathcal{A}_d$ , we have

$$\frac{1}{\bar{h} - \underline{h} + 1} \cdot \sum_{h=\underline{h}}^{\bar{h}} \mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ \frac{\text{REW}_A(I_h, s_1)}{\text{OPT}(I_h)} \right] = 1 - \Theta\left(\frac{1}{p\sqrt{m}}\right).$$

By Equation (56) in the proof of Theorem 2, it suffices to show that there does not exist a mapping  $y_0(\cdot) : \{0, \dots, \bar{h}\} \rightarrow \{0, 1, \dots, m\}$ , such that

$$\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ (h - y_0(s_1))^{\frac{r_1}{r_2}} \mathbf{1}(h \geq y_0(s_1)) + (y_0(s_1) - h) \right]}{p\sqrt{m}} = 0. \quad (70)$$

Recall that under family  $\mathbf{F}$ , any deterministic algorithm can be presented with such a mapping.

We show this by contradiction. Contrary to our claim, suppose there exists a mapping  $y_0(\cdot)$  such that Equation (70) holds. This and the fact that  $(h - y_0(s_1))^{\frac{r_1}{r_2}} \mathbf{1}(h \geq y_0(s_1)) \geq 0$ , imply that

$$\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} \leq 0. \quad (71)$$

Next we consider two cases. In the first case, we assume that

$\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} < 0$  and we reach a contradiction under this assumption. In the second case, assume that  $\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} = 0$  and again reach a contradiction.

*Case 1:*  $\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} < 0$ . Suppose that  $\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} = \eta < 0$ . Then, we show that  $y_0$  does not satisfy Equation (70), which is a contradiction. The left hand side of Equation (70) can be written as

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ (h - y_0(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq y_0(s_1)) + (y_0(s_1) - h) \right]}{p\sqrt{m}} \\ & \geq \lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ (h - y_0(s_1)) \frac{r_1}{r_2} + (y_0(s_1) - h) \right]}{p\sqrt{m}} \\ & = -\frac{r_1}{r_2} \eta + \eta \\ & = -\frac{r_1 - r_2}{r_2} \eta \\ & > 0, \end{aligned}$$

which contradicts Equation (70).

*Case 2:*  $\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h}} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} [y_0(s_1) - h]}{p\sqrt{m}} = 0$ . Similar to the proof of Theorem 2, we compare  $y_0(\cdot)$  and  $y(\cdot)$ , where  $y(s_1) = \frac{s_1}{p}$  for any  $s_1 \in \{0, 1, \dots, h\}$ , and then show a contradiction. Equation (58) in the proof of Theorem 2 has shown that if  $y(s_1) = \frac{s_1}{p}$ , then for any  $h \in [\underline{h}, \bar{h}]$ ,

$$\lim_{m \rightarrow \infty} \frac{\mathbb{E}_{s_1 \sim \text{Bin}(h,p)} \left[ (h - y(s_1)) \frac{r_1}{r_2} \mathbf{1}(h \geq \frac{s_1}{p}) + (y(s_1) - h) \right]}{p\sqrt{m}} > 0. \quad (72)$$

By subtracting Equation (72) from Equation (70), we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h}} \sum_{h=\underline{h}}^{\bar{h}} \left( \mathbb{E}_{\mathbf{s}} \left[ (hp - s_1) \frac{r_1}{r_2} \mathbf{1}(hp \geq s_1) + (s_1 - hp) \right] \right. \\ & \left. - \mathbb{E}_{\mathbf{s}} \left[ (hp - y_0(s_1)p) \frac{r_1}{r_2} \mathbf{1}(hp \geq y_0(s_1)p) + (y_0(s_1)p - hp) \right] \right) > 0. \end{aligned} \quad (73)$$

As in this case, we have  $\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_{\mathbf{s}} [s_1 - y_0(s_1)p]}{p\sqrt{m}} = 0$ , Equation (73) can be written as

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \mathbb{E}_{\mathbf{s}} \left[ (hp - s_1) \frac{r_1}{r_2} \mathbf{1}(hp \geq s_1) - (hp - y_0(s_1)p) \frac{r_1}{r_2} \mathbf{1}(hp \geq y_0(s_1)p) \right] \right) > 0. \quad (74)$$

Using the same set  $\mathcal{D}$  defined in the proof of Theorem 2 and following similar steps, we can show that

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D}^c \cap [h]} \Pr(s_1 = d) \left[ (hp - d) \frac{r_1}{r_2} \mathbf{1}(hp \geq d) - (hp - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(hp \geq y_0(d)p) \right] \right) = 0, \quad (75)$$

Therefore, we have

$$\lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \left( \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d) \left[ (hp - d) \frac{r_1}{r_2} \mathbf{1}(hp \geq d) - (hp - y_0(d)p) \frac{r_1}{r_2} \mathbf{1}(hp \geq y_0(d)p) \right] \right) > 0. \quad (76)$$

By the above inequality and following similar steps in the proof of Theorem 2, we can show that

$$\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d) > 0. \quad (77)$$

We are now ready to show the contradiction using the assumption in this case, i.e.,

$$\lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_s [y_0(s_1) - h]}{p\sqrt{m}} = 0.$$

We have that

$$\begin{aligned} & \lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \frac{\mathbb{E}_s [y_0(s_1) - h]}{p\sqrt{m}} \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \left( \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)(y_0(d)p - hp) + \sum_{d \in \mathcal{D}^c \cap [h]} \Pr(s_1 = d)(y_0(d) - h) \right) \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \left( \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)(y_0(d)p - d + d - hp) + \sum_{d \in \mathcal{D}^c \cap [h]} \Pr(s_1 = d)\left(\frac{d}{p} - h\right) \right) \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \left( \sum_{d \in [h]} (d - hp) + \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)(y_0(d)p - d) \right) \\ &= \lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \mathbb{E}_{d \sim \text{Bin}(h, p)} [d - hp] + \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)(y_0(d)p - d) \\ &= \lim_{m \rightarrow \infty} \frac{1}{p\sqrt{m}} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)(y_0(d)p - d) \\ &\geq \lim_{m \rightarrow \infty} \frac{1}{\bar{h} - \underline{h} + 1} \sum_{h=\underline{h}}^{\bar{h}} \sum_{d \in \mathcal{D} \cap [h]} \Pr(s_1 = d)c_2 \\ &> 0, \end{aligned}$$

The last inequality is from Equation (77). This is a contradiction to the assumption in case 2. Therefore, there does not exist a mapping  $y_0(\cdot)$  with which we can achieve a competitive ratio of  $1 - o(\frac{1}{p\sqrt{m}})$ .

### E.1. Proof of Lemma 15

Let  $P(A)$  be the probability of choosing an algorithm  $A$ . Further, let  $Q(X)$  be the probability of choosing meta input  $X \in \mathcal{X}$ . We start from the right hand side of Equation (68):

$$\begin{aligned} \max_{A \in \mathcal{A}_d} \mathbb{E}_{X \sim Q} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]] &= \sum_{A \in \mathcal{A}_d} P(A) \cdot \max_{A \in \mathcal{A}_d} \mathbb{E}_{X \sim Q} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]] \\ &\geq \sum_{A \in \mathcal{A}_d} P(A) \cdot \mathbb{E}_{X \sim Q} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]] \\ &= \sum_{A \in \mathcal{A}_d} P(A) \cdot \sum_{X \in \mathcal{X}} Q(X) \cdot \mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})] \\ &= \sum_{X \in \mathcal{X}} Q(X) \cdot \sum_{A \in \mathcal{A}_d} P(A) \cdot \mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})] \\ &= \sum_{X \in \mathcal{X}} Q(X) \cdot \mathbb{E}_{A \sim P} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]] \\ &\geq \min_{X \in \mathcal{X}} \mathbb{E}_{A \sim P} [\mathbb{E}_{\tilde{X} \sim D(X)} [C_A(\tilde{X})]], \end{aligned}$$

where the last inequality is the desired result.

## Appendix F: Proof of Theorem 4

We define a family of arriving sequence  $\mathbf{F}_1$  as follows: The input family  $\mathbf{F}_1$  contains all  $(h, \ell)$  such that  $h \in [0, m/p]$  and  $\ell = \frac{10000 \cdot m}{p}$ . For any  $h \in [0, m/p]$ , we then denote  $I_h$  as an ordered random arrival sequence under which  $n_2 \sim \text{Bin}(\ell, 1-p)$  type 2 agents arrive followed by  $n_1 \sim \text{Bin}(h, 1-p)$  type 1 agents. Then, we characterize an upper bound on the CR of any deterministic algorithm under the family  $\mathbf{F}_1$ . Similar to the proof of Theorem 2, given that the online arrival sequences  $I_h$  are all ordered, any deterministic algorithm has to decide about how many type 2 agents they accept provided that they observe  $s_1$  samples from type 1 agents. Put differently, any deterministic algorithm can be represented by a mapping that maps  $s_1$  to the number of type 2 agents it accepts.

We pick two sequences  $I_0$  and  $I_{\tilde{h}}$  inside the family  $\mathbf{F}_1$ , where under  $I_0$ , no type 1 agents arrive and under  $I_{\tilde{h}}$ , the market size for type 1 agents is  $\tilde{h}$ . Here,  $\tilde{h}$  is the smallest integer such that  $\tilde{h}p + \sqrt{\tilde{h} \log(\tilde{h})} \geq 1$ . It is easy to see that  $\tilde{h} \leq 1/p$  and  $\tilde{h} = O(1/p)$ .

For the sequence  $I_0$ ,  $\Pr(s_1 = 0) = 1$  because  $h = 0$ . For the sequence  $I_{\tilde{h}}$ ,  $\Pr(s_1 = 0) \geq 1 - \frac{1}{\tilde{h}^2}$  because by Hoeffding's inequality,

$$\Pr\left(s_1 - \mathbb{E}[s_1] \geq \sqrt{\tilde{h} \log(\tilde{h})}\right) \leq \frac{1}{\tilde{h}^2}. \quad (78)$$

As  $\mathbb{E}[s_1] = \tilde{h}p$ , Equation (78) is equivalent to

$$\Pr\left(s_1 - \tilde{h}p \geq \sqrt{\tilde{h} \log(\tilde{h})}\right) = \Pr(s_1 \geq \tilde{h}p + \sqrt{\tilde{h} \log(\tilde{h})}) \leq \Pr(s_1 \geq 1) \leq \frac{1}{\tilde{h}^2},$$

which implies that

$$\Pr(s_1 = 0) \geq 1 - \frac{1}{\tilde{h}^2}.$$

Now suppose that an algorithm accepts  $m-x$  type 2 agents if the decision-maker observes  $s_1 = 0$ . Then the CR under  $I_0$  is  $\frac{m-x}{m} = 1 - \frac{x}{m}$ . The expected CR under  $I_{\tilde{h}}$  (denoted by  $\text{CR}(I_{\tilde{h}})$ ) is bounded by

$$\begin{aligned} \text{CR}(I_{\tilde{h}}) &= \mathbb{E}[\text{CR}(I_{\tilde{h}})|s_1 = 0] \Pr(s_1 = 0) + \mathbb{E}[\text{CR}(I_{\tilde{h}})|s_1 \geq 1] \Pr(s_1 \geq 1) \\ &\leq \frac{(m-x)r_2 + xr_1}{\min\{\tilde{h}, m\}r_1 + (m-\tilde{h})r_2} + \frac{1}{\tilde{h}^2}, \end{aligned}$$

where we use the fact that  $\Pr(s_1 = 0) \leq 1$  and  $\mathbb{E}[\text{CR}(I_{\tilde{h}})|s_1 \geq 1] \leq 1$ . Therefore, the CR is bounded by

$$\min\left\{1 - \frac{x}{m}, \frac{(m-x)r_2 + xr_1}{\min\{\tilde{h}, m\}r_1 + (m-\tilde{h})r_2} + \frac{1}{\tilde{h}^2}\right\}.$$

To find the optimal  $x$ , we can let the two terms above be equal, and we have

$$x^* = m - \frac{m^2}{m + \min\{\tilde{h}, m\}(1-\alpha)} - m \frac{1}{\tilde{h}^2},$$

and the upper bound of CR can be calculated as

$$1 - \frac{x^*}{m} \leq \frac{m}{m + \min\{\tilde{h}, m\}(1-\alpha)} + \frac{1}{\tilde{h}^2},$$

which is the desired result.

The proof for randomized algorithms is omitted because we can apply Lemma 15 and use the similar method in the proof of Theorem 2 to derive the upper bound among all randomized algorithms.

### Appendix G: Proof of Proposition 3

The proof is similar to the proof of Theorem 1, where there is no capacity constraint during the test period. When there is a capacity constraint (i.e., when the number of resources during the test period is  $m_t = \omega(\sqrt{m})$ ), the number of samples is not longer  $s_1$  and  $s_2$  under the case of  $s_1 + s_2 > m_t$ . This, in turn, influences the estimation of  $\hat{r}_i$ ,  $i \in \{1, 2\}$ , and the probability of the good event that  $\hat{r}_1 > \hat{r}_2$ .

We note that in the proof of Theorem 1, the only place we needed a (lower) bound on the probability of the good event is in the analysis of the third region  $\mathcal{R}_3 = \{(h, \ell) : h \geq \sqrt{m}, \ell \geq \sqrt{m}\}$ . The results for the other two regions are still valid. Hence, in the following, we focus on the third region.

Given that  $h, \ell \geq \sqrt{m}$  in the third region and  $s_1 + s_2 > m_t$ , we split the analysis into two cases:

1.  $s_1 > m_t/2$ ,  $s_2 > m_t/2$ . In this case, the decision-maker accepts  $m_t/2$  each type agents. Then,  $\hat{r}_i = \frac{\text{Bin}(m_t/2, r_i)}{m_t/2}$ ,  $i \in \{1, 2\}$ . Then, we replace  $s_i$  by  $m_t/2$  in the proof of Lemma 7 and as  $m_t = \omega(\sqrt{m})$ , we can get the same asymptotic result.
2.  $s_1 < m_t/2$ ,  $s_2 > m_t/2$  or  $s_1 > m_t/2$ ,  $s_2 < m_t/2$ . Without loss of generality, we analyze the first case (i.e.,  $s_1 < m_t/2$ ,  $s_2 > m_t/2$ ). In this case, we accept  $s_1$  type 1 agents and  $m_t - s_1$  type 2 agents. Then,  $\hat{r}_1$  remains the same and  $\hat{r}_2 = \frac{\text{Bin}(m_t - s_1, r_2)}{m_t - s_1}$ . Then, we replace  $s_2$  by  $m_t - s_1$  in the proof of Lemma 7 and as  $m_t = \omega(\sqrt{m})$ , we can also get the same asymptotic result.

### Appendix H: Proof of Proposition 4

Consider the instance  $I$  which contains  $\ell = \frac{10000 \cdot m}{p \cdot r_2}$  type 2 agents and 0 type 1 agent. Then, in the online arrival sequence, with probability 1, the number of realized rewards of 1 is larger than  $m$ . Therefore,  $\text{OPT-RE}(\mathcal{R}, I) = m$  almost surely.

As the realization of each arriving agent is independent and follows  $\text{Ber}(r_2)$ , any algorithm cannot predict the realization of the arriving agent. If one accepts an agent, then the expectation of reward we get is  $1 \cdot \Pr(\text{Ber}(r_2) = 1) + 0 \cdot \Pr(\text{Ber}(r_2) = 0) = r_2$ . The best algorithm can accept  $m$  type 2 agents due to the resource constraint. Therefore, the realized competitive ratio is at most  $\frac{mr_2}{m} = r_2$ .

### Appendix I: Proof of Proposition 5

We first consider region  $\mathcal{R}_1$ . From Lemma 2, CR is  $1 - \Theta(1/\sqrt{m})$  which does not depend on  $p$  in all cases except  $\{n_2 < m - n_1\}$  or  $\{n_2 \geq m - n_1\} \cap \{n_1 \leq \frac{1-p_1}{p_1} s_1\}$ . Now, suppose that we apply the modified version of Algorithm 1. This means we give protection level  $\min\{\frac{1-p_1}{p_1} s_1, m\}$  to type 1 agent when the algorithm believes type 1 is the high-reward type and give protection level  $\min\{\frac{1-p_2}{p_2} s_2, m\}$  to type 2 agent when the algorithm believes type 2 is the high-reward type. As the CR does not depend on  $p$ , in all cases except  $\{n_2 < m - n_1\}$  or  $\{n_2 \geq m - n_1\} \cap \{n_1 \leq \frac{1-p_1}{p_1} s_1\}$ , the CR is still  $1 - \Theta(1/\sqrt{m})$ .

For the case of  $n_2 < m - n_1$ , we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &= \frac{\min\left\{n_2, \left(m - \frac{1-p_1}{p_1} s_1\right)^+\right\} \cdot r_2 + n_1 r_1}{n_2 r_2 + n_1 r_1} \\ &\geq \frac{\min\left\{n_2, \left(m - \frac{1-p_1}{p_1} s_1\right)^+\right\}}{n_2} = \min\left\{1, \frac{\left(m - \frac{1-p_1}{p_1} s_1\right)^+}{n_2}\right\} \end{aligned}$$

$$= \frac{\left(m - \frac{1-p_1}{p_1} s_1\right)^+}{n_2} > \frac{\left(m - \frac{1-p_1}{p_1} s_1\right)^+}{m - n_1} \geq \frac{\left(m - \frac{1-p_1}{p_1} h\right)^+}{m} \geq \left(1 - \frac{1-p_1}{p_1} \frac{1}{\sqrt{m}}\right)^+,$$

where in the last inequality holds because in region  $\mathcal{R}_1$ ,  $h < \sqrt{m}$ .

For the case  $\{n_2 \geq m - n_1\} \cap \{n_1 \leq \frac{1-p_1}{p_1} s_1\}$ , we have

$$\begin{aligned} \frac{\text{REW}_A(\mathbf{n}, \boldsymbol{\rho}; \mathcal{G})}{\text{OPT}(\mathbf{n})} &= \frac{\min\{n_2, (m - \frac{1-p_1}{p_1} s_1)^+\} r_2 + n_1 r_1}{(m - n_1) r_2 + n_1 r_1} \\ &= \frac{\left(m - \frac{1-p_1}{p_1} s_1\right)^+ r_2 + n_1 r_1}{(m - n_1) r_2 + n_1 r_1} \geq \frac{\left(m - \frac{1-p_1}{p_1} s_1\right)^+}{m - n_1} \geq \frac{\left(m - \frac{1-p_1}{p_1} h\right)^+}{m} \geq \left(1 - \frac{1-p_1}{p_1} \frac{1}{\sqrt{m}}\right)^+. \end{aligned}$$

For any other cases, we still have the CR is  $1 - \Theta(1/\sqrt{m})$ . Totally, we have the modified version of Algorithm 1 in  $\mathcal{R}_1$  has a CR of  $1 - \Theta(1/(p_1 \sqrt{m}))$ .

Next, from Lemma 3, in region  $\mathcal{R}_2$ , CR is  $1 - \Theta(1/\sqrt{m})$ , which does not depend on  $p$ . Therefore, the modified version also achieves a CR of  $1 - \Theta(1/\sqrt{m})$ .

Finally, from Lemma 4, in region  $\mathcal{R}_3$ , for  $m \geq m_1$ ,  $\overline{\text{CR}}_3 = 1 - \Theta(1/(p^2 m))$ ; and for  $m < m_1$ ,  $\overline{\text{CR}}_3 = 1 - \Theta(1/\sqrt{m})$ . Again, if we apply the modified version of Algorithm 1, for  $m < m_1$ , the CR is still  $1 - \Theta(1/\sqrt{m})$ . Otherwise, from the proof of Lemma 7, we have

$$\overline{\text{CR}}_3 \geq \left(1 - \frac{1}{(\sqrt{m} p_1 - m^{1/4})^2}\right) \left(1 - \frac{1}{(\sqrt{m} p_2 - m^{1/4})^2}\right) = 1 - \Theta\left(\frac{1}{\min\{p_1, p_2\}^2 m}\right).$$

To summarize, we have the modified version of Algorithm 1 has a CR of

$$1 - \Theta\left(\max\left\{\frac{1}{p_1 \sqrt{m}}, \frac{1}{\min\{p_1, p_2\}^2 m}\right\}\right).$$

## Appendix J: Algorithm for Sampling Probability $p = O(1/\sqrt{m})$

---

### Algorithm 3 Algorithm 3 for $p = O(1/\sqrt{m})$

---

**Input:** The number of resources  $m$ .

1. With probability  $1/2$ , we only accept type 1 agents. With probability  $1/2$ , we only accept type 2 agents.
- 

**THEOREM 5 (Competitive Ratio of Algorithm 3).** *Consider the model presented in Section 2, where the expected rewards of  $r_i$ ,  $i \in \{1, 2\}$  is unknown to the decision-maker. For  $p = O(1/\sqrt{m})$  and for any  $m > 0$ , the CR of Algorithm 3 is  $1/2$ .*

*Proof of Theorem 5* With probability  $0.5$ , we reject all type 2 agents and accept all type 1 agents, and with probability  $0.5$ , we reject all type 1 agents and accept all type 2 agents. Then, we have

$$\frac{\mathbb{E}[\text{REW}_A(I)]}{\text{OPT}(I)} = \frac{\frac{1}{2} \min\{n_1, m\} r_1 + \frac{1}{2} \min\{n_2, m\} r_2}{\min\{n_1, m\} r_1 + \min\{n_2, (m - n_1)^+\} r_2} \geq \frac{1}{2},$$

where the last inequality is due to  $\min\{n_2, (m - n_1)^+\} \leq \min\{n_2, m\}$ .

□



## Appendix K: Partially Known Sampling Probability

In our setting, we assume that the sampling probability  $p$  is known to the decision-maker, and we show that under this assumption, Algorithm 1 obtains an asymptotically optimal CR. However, as we show in Section 3, obtaining the same result is not possible when  $p$  is completely unknown, justifying our assumption. Nonetheless, here we would like to study a setting in which the decision-maker (before the test period starts) obtains an estimate of the sampling probability, denoted  $\hat{p}$ , where  $\hat{p} \in [p(1 - \delta), p(1 + \delta)]$  for some  $\delta \in (0, 1)$ . Here,  $p$  is the true (unknown) sampling probability and  $\delta$  measures how accurate the estimate is. Such an estimate can be obtained using previous allocation periods in which the same outreach program is used.

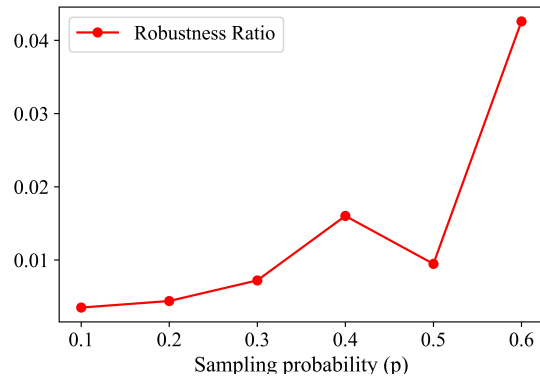
For this setting, let us consider a variation of Algorithm 1 under which the algorithm pretends the estimated sampling probability to be the true probability. That is, when the algorithm protects type 1 agents (i.e.,  $\hat{r}_1 > \hat{r}_2$ ), the algorithm gives a protection level of  $\min\{m, s_1 \frac{1-\hat{p}}{\hat{p}}\}$  to type 1 agent. Otherwise, the algorithm gives  $\min\{m, s_2 \frac{1-\hat{p}}{\hat{p}}\}$  protection level to type 2 agent. The following example shows that Algorithm 1 is quite robust to the estimation error in the sampling probability under adversarial arrivals. Recall that in Section 6, we show that Algorithm 1 with noisy  $p$  performs extremely good under stochastic arrivals.

EXAMPLE 4. Let  $\text{CR}(p; \hat{p})$  be the worst case CR of Algorithm 1 with input  $\hat{p}$  and the true sampling probability  $p$ . Then, we measure the robustness of Algorithm 1 using the following measure that we refer to as *robustness ratio*:

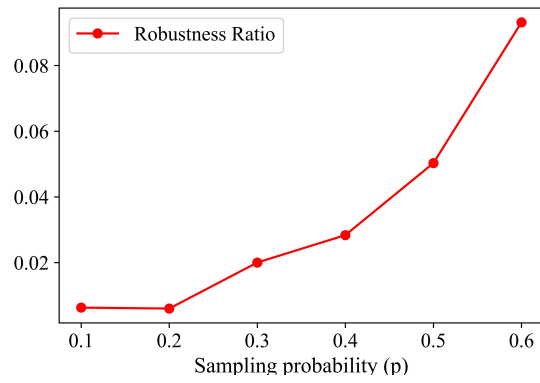
$$\sup_{\hat{p} \in [p(1-\delta), p(1+\delta)]} \frac{\text{CR}(p; p) - \text{CR}(p; \hat{p})}{\text{CR}(p; p)}.$$

Observe that the smaller the robustness ratio, the better. The CR is calculated by taking the minimum value among 5000 generated arrival instances.

Figures 13 and 15 present the robustness ratio versus  $p$  with  $\delta \in \{0.1, 0.2\}$ ,  $r_1 = 0.9$ , and  $r_2 = 0.5$  for the case of  $m = 30$  and  $m = 200$ , respectively. The figures show the robustness ratio mainly depends on the estimation error  $\delta$ . When  $\delta = 0.1$ , the robustness ratio is in the range of 1% – 6%. That is, the CR can drop by at most 6% when we have 10% estimation error in  $p$ . Similarly, when with  $\delta = 0.2$ , the robustness ratio is in the range of 1 – 12%. These results confirm the robustness of Algorithm 1 to the estimation error in  $p$ .

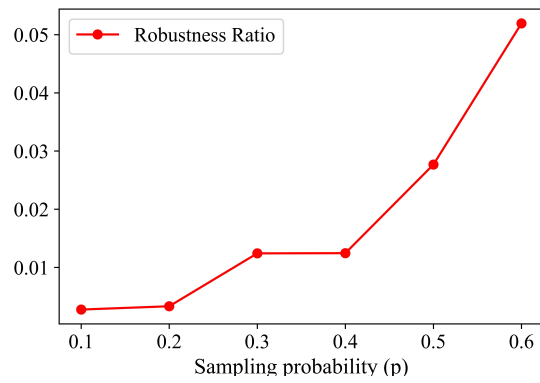


**Figure 11**  $\delta = 0.1, m = 30$

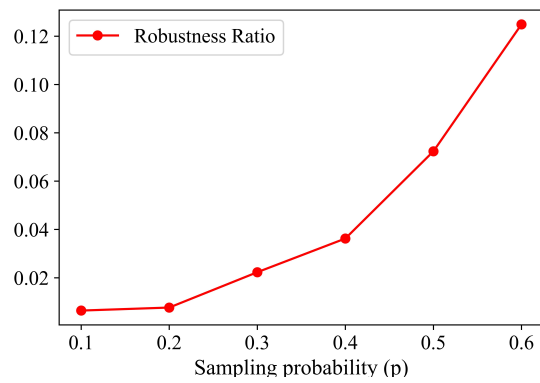


**Figure 12**  $\delta = 0.2, m = 30$

**Figure 13** Robustness of Algorithm 1 to the estimation error in  $p$  for  $m = 30$ . Here,  $r_1 = 0.9$ , and  $r_2 = 0.5$ .



**Figure 14**  $\delta = 0.1, m = 200$



**Figure 15**  $\delta = 0.2, m = 200$

**Figure 16** Robustness ratio of Algorithm 1 in  $p$  for  $m = 200$ . Here,  $r_1 = 0.9$ , and  $r_2 = 0.5$ .