Algorithms for Online Optimization Problems in Operations

by

Jiashuo Jiang

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Stern School of Business

New York University

September, 2022

Professor Jiawei Zhang and Professor Will Ma

To my families, to everyone I love, and everyone loves me.

Acknowledgements

I would like to thank so many people. Without their help and support, this thesis would never be possible to be finished.

I am forever beholden to my advisors, Prof. Jiawei Zhang and Prof. Will Ma. I definitely cannot achieve what I own now without their endless guidance and support. I can never express enough my respect and gratitude for their amazing guidance, which always help me get through the hard time and out of the dilemma. I start doing research with Jiawei from the reading group. Jiawei has been so patient to give me hand-by-hand and well-rounded instructions over so many things and teach me from elementary stuff. He is also so knowledgeable and always point out the right way to go. He has led me into the broad literature on online optimization and from then, I decided to devote my PhD study to this area. In the following years, Jiawei not only cares about me as a person, but also teaches me everything to grow into an independent researcher. He has always inspired and motivated me with his curiosity over new techniques, passion over new research ideas, and his tastes over what is a good research. He helps me out with his intelligence so many times and most importantly, he establishes a role model of an excellent scholar, not only for me, but also for other young scholars. I will never learn how to do research without Jiawei's guidance and his great personality will motivate me to achieve more in my future path. The first time I met Will is in an Informs session and I was so impressed by his intelligence. He is also so friendly to young students. We start doing research together since then and he is so patient over me and cares me a lot. I am always surprised at how smart he is, yet still very approachable

and patient. He has spend so many hours giving me very detailed instructions over many things. Will is a great teacher. He always taught me with a very simple example which is very easy to understand but still provides enough insights for me to understand the whole pattern. He is so intelligent and I can never learn enough from him. Besides academy, Will also cares about me in terms of my life. He understands the ups and downs of me and he offers me help even before I ask for it. I have always been impressed by his curiosity, intelligence and his genuine attitude towards research, which greatly motivates me to devote more into research and deliver more great results in the future. I can never express enough how I benefit from Jiawei and Will. They have used their knowledge, intelligence, and personality to help me grow from an ignorant student into an independent research, and motivate me to become a better scholar in the future. I am so proud to be advised by them. I can never express enough how much I thank them. I am forever indebted to them.

I would like to thank Prof. Xi Chen, Prof. Srikanth Jagabathula and Prof. Zhengyuan Zhou for agreeing to serve on my committee. I thank them for their valuable instructions through my PhD study. I would also like to Prof. Ilan Lobel for writing reference letters for me. The conversation with Ilan has benefited my own research a lot. I also would like to thank other faculty members from NYU Stern, including Prof. Mor Armony, Prof. Michael Pinedo, Prof. Joshua Reed, Prof. Wenqiang Xiao, Prof. Elynn Chen and Prof. Divya Singhvi for the courses they provide and the advice they have given to my research. My special thanks go to my undergraduate mentor, Prof. Zaiwen Wen, who led me to the Operations field and referred me to NYU, such a wonderful place, to do my graduate study.

I greatly thank Xiaocheng Li, Shixin Wang, and Boxiao Chen for the collaborations and for the help I have received from them. Life at New York City is not easy, especially for PhD students. I would like to thank my other friends at NYU, including Ziran Liu, Haotian Song, Yichen Zhang, Xintian Han, Sen Yang, Weichi Yao, Xinyi Zhao, Zhuoyi Yang, Dmitry Mitrofanov, He Li, and Sandeep Chitla, Xiaole Liu, Fanglin Chen, Pan Li, Chenshuo Sun and Xinyu Zhang. I also would

like to thank my other friends outside NYU, including Ruohan Zhan, Bangrui Chen, Yilun Chen, Shi Dong, Yaqi Duan, Yichun Hu, Yue Hu, Xiao Lei, Heyuan Liu, Xin Qian, Liang Xue, Wenzhao Ban, Luhao Zhang and so many others. I also want to thank my manager Michael Bloem at Amazon, who has helped and inspired me a lot. My life would not be so enjoyable without their accompanies.

Last but most importantly, I would like to thank my parents, Chenghai Jiang and Fuhong Zhang, for their permanent love and unconditional support. Life has been so difficult for me during the Covid period. It is their love supporting me to get through all of these and finally finish my PhD successfully. I thank my families so much for their unconditional backup since I was born. I cannot achieve anything without them. I dedicate this dissertation to them.

All in all, I feel so fortunate to have so many wonderful memories, to receive so much help, and have so many excellent experiences with so many people during my time at NYU. All of these motivate me to become a better person, to achieve more in the future, and to finally give back to the communities and society.

Abstract

With the development of technologies and the booming of the online economy, it is becoming more and more important for companies to make their operational decisions in an online manner. This requires us to solve the online optimization problem where the parameters are arriving one by one, and the decisions need to be made in real-time. In this thesis, we will address the issue of how to design efficient algorithms for online optimization problems rooted from operations applications and how to obtain strong theoretical guarantees for the algorithms.

In the first part, we consider designing online algorithms with strong competitive ratio guarantees, which measure the relative difference of the online algorithms with respect to the offline benchmarks. Such problems are also known as prophet inequalities in the literature. we seek tight guarantees, which not only pursue the theoretical limit of the performance of any online algorithms, but also lead to new algorithms that are optimal. The key techniques root in analysis of the linear programming formulations of the online problems. The specific structures of the linear programming formulations are revealed and optimal solutions are found in closed formulations, the first time in the literature.

In the second part, we consider designing online algorithms with strong regret guarantees, which measure the additive difference of the online algorithms with respect to the benchmarks. Different information regimes are considered, including given or unknown distributions under a stationary or non-stationary environment. The online algorithms are developed for all these settings. The main techniques root in online convex optimization theories, online learning theories,

and primal-dual analysis of the optimization problems. This thesis also discusses how the online algorithms here can be applied to solve a wide range of real-life problems arising from the operational fields, such as supply chain management, inventory management, revenue management, online advertising, and mechanism design.

Contents

1 Introduction

The problem of making decisions in an online environment, where the system parameters arrive one by one and the decision have to be made sequentially, lies at the center of operations research and optimization. Especially with the development of technologies and the booming of the online economy, it is becoming more and more important for companies to make their operational decisions in a sequential way by solving an online optimization problem. In this thesis, we aim to develop near-optimal algorithms for various online optimization problems that arise from the operations field of applications. Moreover, in addition to practical performance, we seek to obtain a strong theoretical guarantees for our algorithms that provide robustness justifications for our algorithms. We measure the performance of our algorithms in two ways, the competitive ratio and the regret, which we illustrate in the following parts.

1.1 Results for Competitive Analysis

In the first part, we aim to derive near-optimal algorithms for online optimization problems with a strong *competitive ratio* guarantee, which measures the *relative* difference of the performance of the online algorithm against the offline benchmark. Such problems are also known as *prophet inequalities* (Jiang et al., 2022a,b). A keystone ingredient for algorithm design for prophet inequalites has been to start with a *Linear Programming (LP) relaxation*, which is also known as the ex-ante, fluid, or deterministic relaxations. The LP relaxation directly generalizes prophet

inequality to a general online resource allocation problem with multiple resources, which arises in many domains such as online advertising, healthcare operations, and revenue management. Our first new result is to derive the tight guarantees with respect to the LP relaxation, for two important prophet inequalities, namely the multi-unit prophet inequality and the knapsack prophet inequality, which are often used as crucial subroutines in mechanism design and general online resource allocation problems. We then go beyond the LP relaxation. Our second new result is a new LP framework, which enables us a unified way to derive guarantees for both the LP relaxation and the prophet directly.

1.1.1. New Result I: Tight LP-based Guarantees. We consider multi-unit prophet inequalities and the knapsack prophet inequality. In these problems, there is a single resource with an infinitely-divisible capacity normalized to 1. A sequence of queries is observed, each with a reward and a size. Upon observation, a query must be immediately either served, claiming its reward and spending capacity equal to its size; or rejected, which is the only option if the remaining capacity is less than its size. The (size, reward) pair for each query is initially unknown, but drawn independently from non-identical distributions that are given in advance. The objective is to maximize the total reward collected in expectation. The special case of k-unit prophet inequalities restrains all query sizes to be $1/k$, while the general knapsack prophet inequality allows the size of each query to be arbitrary between 0 and 1.

We derive the *best-possible* guarantees with respect to the LP relaxation for both the multi-unit prophet inequalities and the knapsack prophet inequality. Specifically, the *k-unit* problem originated from Hajiaghayi et al. (2007), and a celebrated guarantee of $1 - 1/\sqrt{k+3}$ was derived by Alaei (2011) through a "Magician's problem". These guarantees have been recently improved in Wang et al. (2018); Chawla et al. (2020), and in this paper we fully resolve (see Table 1.1) the question of k-unit prophet inequalities relative to the LP. Meanwhile, for the knapsack problem, we derive the tight guarantee of $\frac{1}{3+e^{-2}} \approx 0.319$ which improves earlier results of Dutting et al. (2020) and Alaei et al. (2013). We emphasize that our results can be *directly applied* to improve

the guarantees in *all of the papers cited thus far*. The tightness of our results also characterizes the *limits* of any approach that seeks approximately-optimal algorithms through the LP relaxation. Finally, in the case of unit-density online knapsack, where the random size and reward of a query are always identical, our analysis for knapsack can be improved to obtain an improved guarantee 0.3557.

Table 1.1: Our tight ratios for multi-unit prophet inequalities. The previous lower bounds are obtained as the maximum of the ratios in Alaei (2011); Chawla et al. (2020). The previously best-known upper bounds were 1/2 for $k = 1$ and $1 - e^{-k}k^k/k!$ for $k > 1$, with the latter inherited from the "correlation gap" in the i.i.d. special case (see Yan, 2011).

value of k	1	2	3	4	5	6	7	8
Existing lower bounds	0.5000	0.5859	0.6309	0.6605	0.6821	0.6989	0.7125	0.7240
Our tight ratios	**0.5000**	**0.6148**	**0.6741**	**0.7120**	**0.7389**	**0.7593**	**0.7754**	**0.7887**
Existing upper bounds	0.5000	0.7293	0.7760	0.8046	0.8245	0.8394	0.8510	0.8604

All in all, our results imply tight (non-greedy) Online Contention Resolution Schemes (OCRS) for k-uniform matroids and the knapsack polytope, respectively, which has further implications.

Comparison with literature. k-unit prophet inequalities have been analyzed using LP's before in Alaei et al. (2012), who formulate a primal LP encoding the adversary's problem of minimizing an online algorithm's optimal dynamic programming value. They then use an auxiliary "Magician's problem," analyzed through a "sand/barrier" process, to construct a feasible dual solution with guarantee $\gamma = 1 - \frac{1}{\sqrt{k+3}}$. By contrast, we directly formulate the multi-unit prophet inequality problem as a new LP and find the optimal solution in closed formulations. Our LP dual along with complementary slackness allows us to establish the structure of the optimal policies for k-unit prophet inequalities, showing that it indeed corresponds to a γ-Conservative Magician in Alaei et al. (2012). However, in our case γ is set to a value dependent on on the distributions, which we show is always at least γ_k^*, and strictly greater than $1 - \frac{1}{\sqrt{k+3}}$ for all $k > 1$.

The values of γ_k^* we derive have previously appeared in Wang et al. (2018) through the stochastic analysis of a "reflecting" Poisson process. Our work differs by establishing *optimality* for these values γ_k^*, as the solutions to a sequence of optimization problems from our framework. More-

over, their paper assumes Poisson arrivals to begin with, while we allow arbitrary arrivals and show the limiting Poisson case to be the worst case.

We should note that classically in the k-unit prophet inequality problem, the goal is to compute the worst-case performance of an online algorithm, who sequentially observes independent draws from known distributions and can accept k of them, and compare instead to a *prophet*, whose performance is the expected sum of the k highest realizations. The prophet's performance is upper-bounded by the LP relaxation, so our guarantees that are tight relative to the LP also imply the *best-known prophet inequalities to date* for all $k > 1$. We do give an example that demonstrates this guarantee to be "almost" tight even when comparing to the weaker prophet benchmark. To be specific, we provide an upper bound of 0.6269 relative to the prophet when $k = 2$ (Proposition 2.5). Since $\gamma_2^* \approx 0.6148$, this shows that not much improvement beyond γ_k^* is possible relative to the prophet when $k = 2$.

To our knowledge, our analysis for knapsack differs from all existing ones for knapsack in an online setting (Dutting et al., 2020; Stein et al., 2020; Feldman et al., 2021) by eschewing the need to split queries into "large" vs. "small" based on their size (usually, whether their size is greater than 1/2). In fact, any algorithm that *packs large and small queries separately* is limited to $\gamma \leqslant 0.25$ in our problem (Theorem 2.8), whereas our tight guarantee is $\gamma = \frac{1}{3+e^{-2}} \approx 0.319$.

Another related setting for knapsack is the online stochastic generalized assignment problem of Alaei et al. (2013), for which the authors establish a guarantee of $1 - \frac{1}{\sqrt{k}}$, when each query can realize a random size that is at most $1/k$. They eliminate the possibility of "large" queries by imposing k to be at least 2, showing that a constant-factor guarantee is impossible when $k = 1$. Although our problem can handle random sizes, we need to assume that size is observed *before* the algorithm makes a decision, whereas in their problem size is randomly realized *after* the algorithm decides to serve a query. This distinction allows our problem to have a constant-factor guarantee that holds even when queries can have size 1.

1.1.2. New Result II: A Unified Framework to Go Beyond LP-based Guarantees. We fur-

ther consider the multi-unit prophet inequalities and we aim to derive the tight guarantees even relative to the prophet. In addition to dynamic online policies where the decision depends both on the realization of the current query's value and the remaining capacity, we also study simpler but more restrictive *static* threshold policies (Hajiaghayi et al., 2007), where the decision is given by comparing the realized reward of the query against a threshold and the threshold is fixed for each query, regardless of the remaining capacity. A further restriction of *oblivious* static threshold policy was introduced in Chawla et al. (2020), where the threshold depends only on the probability mass of the queries' reward on the support, instead of the support itself.

We develop a general framework to study the performance guarantees of general online policies, as well as (oblivious) static threshold policies, with respect to both the prophet benchmark and the Ex-Ante benchmark, in both the IID and non-IID settings. We denote by I the problem instance and $\mathcal{I}_{k,T}^{\text{IID}}$ (resp, $\mathcal{I}_{k,T}$) the collection of all IID (resp. non-IID) problem instances with T queries and k slots. We also denote by $\mathsf{DP}(I)$ the value of optimal dynamic programming, $\mathsf{ST}(I)$ the value of the optimal non-oblivious threshold policy, $\mathsf{OST}(I)$ the value of the optimal oblivious threshold policy, $\mathsf{Proph}(I)$ the value of the prophet benchmark, and $\mathsf{ExAnte}(I)$ the value of the Ex-Ante benchmark, on the problem instance I.

We introduce a new LP framework, which explicitly computes the *tight* guarantee of an online policy versus both the prophet and the Ex-Ante benchmark. The new dual LP can be interpreted as a "Type Coverage" problem, which we illustrate in Section 3.2. Our framework shows that the guarantee under the worst-case instance in the original problem is equivalent to the guarantee under the worst-case instance for the Type Coverage problem. If the original prophet problem restricted to an *oblivious* static threshold policy, then the corresponding dual restriction is to have *any* static threshold policy. Meanwhile, if the original problem allowed non-oblivious static thresholds, then the dual allows *randomized* static thresholds (which can achieve a strictly greater guarantee for the Type Coverage problem, as we later show). These correspondences in the restrictions, which are not a priori obvious, are consequences of our framework and formalized

in Section 3.2.

Comparison vs. Magician and OCRS problems. Magician (Alaei, 2014) and OCRS (Feldman et al., 2021) are existing auxiliary problems used to derive Ex-Ante prophet inequalities. In these problems, each query has an *active* probability and must be guaranteed to be accepted with a probability at least θ conditional on being active, by the online policy. The goal is to achieve the maximum θ. It will be shown in Chapter 2 that the optimal guarantees for the k-unit Magician/OCRS problems are actually the same, leading to tight Ex-Ante prophet inequalities.

Although ex-ante prophet inequalities are stronger and hence imply non-ex-ante prophet inequalities, no analogue to Magician/OCRS that can analyze *tight non-ex-ante prophet inequalities* has been known, until now. Our new Type Coverage problem essentially generalizes the Magician/OCRS problems to allow for queries to take *non-binary* states, which leads to a formulation of tight guarantees for non-ex-ante prophet inequalities. Our Type Coverage problem is also more general in that it can recover the Magician/OCRS problems in the appropriate setting. We now describe the other new results and simplifications it achieves.

(Oblivious) Static Thresholds in Non-IID Setting. First, for OST policies in the general non-IID setting, we recover the following positive result, which has been previously established in Chawla et al. (2020):

$$\inf_{I \in \mathcal{I}_{k,T}} \frac{\mathsf{OST}(I)}{\mathsf{ExAnte}(I)} \geqslant \mathrm{BernOpt}(k, T). \tag{1.1}$$

In (1.1), $\mathrm{BernOpt}(k, T)$ is an optimization problem over $T - 1$ Bernoulli random variables that tweaks what Chawla et al. (2020) derived, with the formal specification of $\mathrm{BernOpt}(k, T)$ found in Theorem 3.19. We then show this specification to be *tight* for oblivious static thresholds:

$$\inf_{I \in \mathcal{I}_{k,T}} \frac{\mathsf{OST}(I)}{\mathsf{Proph}(I)} \leqslant \mathrm{BernOpt}(k, T), \tag{1.2}$$

even when competing against the weaker prophet benchmark. We derive this upper bound with-

out explicitly constructing any counterexamples. Instead, we show that when the online policy is restricted to OST, the value of the dual Type Coverage problem cannot exceed a quantity (exact formulation given in (3.14)) which can be shown to be equivalent to $\mathrm{BernOpt}(k, T)$.

We then show that even non-oblivious static threshold policies cannot do better than the method of Chawla et al. (2020) in the worst case, although there can be a separation (see Lemma 3.19.1) between OST and ST for a given instance. To do so, we invoke the result of Chawla et al. (2020) that the worst-case instance is achieved when the Bernoulli random variables have equal probabilities. In this case, we show through a fairly technical argument (Theorem 3.20) that randomized static thresholds (corresponding to ST) cannot do better than a deterministic one (corresponding to OST) in the dual Type Coverage problem. In this way, we show that the ratios ST/Proph, ST/ExAnte, OST/Proph, and OST/ExAnte are all equivalent in the worst case, for general non-IID distributions. This equivalence is new to the literature.

(Oblivious) Static Thresholds in IID Setting. When the distributions are IID, we show that

$$\inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{OST}(I)}{\mathsf{ExAnte}(I)} = \inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{DP}(I)}{\mathsf{ExAnte}(I)} = \inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{OST}(I)}{\mathsf{Proph}(I)} = \frac{\mathbb{E}[\min\{\mathrm{Bin}(T, k/T), k\}]}{k}. \tag{1.3}$$

That is, unlike the non-IID setting, even the optimal dynamic policy DP cannot do any better than static thresholds when compared to the Ex-Ante benchmark. We then establish a stronger equivalence between oblivious and non-oblivious static thresholds in the IID setting. Recall that in the non-IID setting, the equivalence was only true in the worst case, i.e., only after taking an infimum over problem instance I were the ratios ST/Proph, ST/ExAnte, OST/Proph, OST/ExAnte equal. In the IID setting, we show that ST is no better than OST with respect to both Proph and ExAnte, on *every* type distribution. As a result, OST's are *instance-optimal* for static threshold policies in the IID setting, and hence $\inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{ST}(I)}{\mathsf{Proph}(I)}$ is also no better than (1.3).

These results are not new in that both the bounds $\inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{OST}(I)}{\mathsf{ExAnte}(I)} \geqslant \frac{\mathbb{E}[\min\{\mathrm{Bin}(T,k/T),k\}]}{k}$ and $\inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{DP}(I)}{\mathsf{ExAnte}(I)} \leqslant \frac{\mathbb{E}[\min\{\mathrm{Bin}(T,k/T),k\}]}{k}$ are folklore, and corresponding upper bounds on $\inf_{I \in \mathcal{I}^{\mathrm{IID}}_{k,T}} \frac{\mathsf{ST}(I)}{\mathsf{Proph}(I)}$

can also be constructed. However, these upper bounds for static threshold policies relative to the weaker prophet benchmark require a complicated family of examples with messy calculations, whereas our framework elegantly establishes the tightness of the ratio $\frac{\mathbb{E}[\min\{\mathrm{Bin}(T,k/T),k\}]}{k}$ without needing to construct any counterexamples.

General Dynamic Policies in IID Setting. When $k = 1$, we obtain a simplified derivation of the classical 0.745 guarantee for DP/Proph under the IID setting (Hill and Kertz, 1982; Correa et al., 2017; Liu et al., 2021). Under the IID setting, the dual problem can be further simplified as a semi-infinite LP by noting that $p_{t,j}$ equals some value p_j across all queries $t \in [T]$. The key step of our derivation is that the simplified semi-infinite LP admits a closed-form optimal solution, which helps us to establish the worst case as $T \to \infty$ and obtain the value of the optimal guarantee.

1.2 Results for Regret Analysis

In the second part, we aim to derive near-optimal algorithms for online optimization problems with a strong *regret* guarantee, which measures the *additive* difference of the performance of the online algorithm against the offline benchmark. We develop algorithms for a variety of online optimization problems with a wide range of applications. Our algorithms not only enjoy strong regret bound, but also are easy to implement.

1.2.1. New Result I: Primal-Dual Algorithm for Non-stationary Online Stochastic Optimization. We consider a general online stochastic optimization problem with multiple budget constraints over a horizon of finite time periods. In each time period, a reward function and multiple cost functions are revealed, and the decision maker needs to specify an action from a convex and compact action set to collect the reward and consume the budgets. Each cost function corresponds to the consumption of one budget constraint. The reward function and the cost functions of each time period are drawn from an unknown distribution, which is non-stationary across

time. The objective of the decision maker is to maximize the cumulative reward subject to the budget constraints. We consider two settings: (i) a data-driven setting where the true distribution is unknown but a prior estimate (possibly inaccurate) is available; (ii) an uninformative setting where the true distribution is completely unknown.

For the data-driven setting, we propose a Wasserstein-based deviation measure, *Wasserstein-based deviation budget* (WBDB), to quantify the deviation of the prior estimate from the true distribution. We illustrate the sharpness of WBDB by showing that if the deviation budget is linear in the total periods, sublinear regret could not be achieved by any admissible policy. Next, we develop a new *Informative Gradient Descent* with prior estimate (IGDP) algorithm, which adaptively combines the prior distribution knowledge into an update in the dual space. We show that the IGDP algorithm achieves the *first* optimal regret upper bound.

For the uninformative setting, we modify the WBDB by replacing the prior estimate of each distribution with their uniform mixture distribution to propose a new measure called *Wasserstein-based non-stationarity measure* (WBNB). We propose *Uninformative Gradient Descent* Algorithm (UGD) as a natural reduction of the IGD algorithm in the uninformative setting. We prove that UGD algorithm achieves a regret bound of optimal order.

Comparison with literature. Our IGDP algorithm is motivated by the traditional online gradient descent (OGD) algorithm (Hazan, 2016). The OGD algorithm applies a linear update rule according to the gradient information at the current period and has been shown to work well in the stationary setting, even when the distribution is unknown (Lu et al., 2020; Sun et al., 2020; Li et al., 2020). The OGD type algorithm has also been developed under stationary setting with unknown distributions in Agrawal and Devanur (2014) by assuming further stronger conditions on the dual optimal solution, and with known distributions for service level problems (Jiang et al., 2019). However, the update in OGD for each time period only involves information gathered up to the current time period, but for the non-stationary setting, we also need to take advantage of the prior estimates of the future time periods. Specifically, based on a primal-dual convex re-

laxation of the underlying offline problem, we obtain a prescribed allocation of the budgets over the entire horizon based on the prior estimates. Then, the IGDP algorithm uses this allocation to adjust the gradient descent direction. This idea is new for the related literature in that the IGDP descent direction at each period does not simply come from the historical observations, but it is also informed by the distribution knowledge of the entire horizon.

A few recent works also study similar online decision making problems in a non-stationary environment. Devanur et al. (2019) study the case of known distribution and obtain a $1-O(1/\sqrt{c})$ competitive ratio, where c denotes the minimal capacity of the budget constraints. It remains unclear how to generalize the method in (Devanur et al., 2019) to the setting where the distribution knowledge is inaccurate or absent. A line of works (Vera and Banerjee, 2020; Banerjee and Freund, 2020a,b) study the known distribution setting with an additional assumption that the underlying distribution takes a finite support. These works develop algorithms that achieve bounded regret that bears no dependency on the total periods. Compared to this stream of literature, the main results of our paper do not assume the finite supportedness. In addition, when the underlying distribution is finite, we can extend the previous algorithm and analysis for the data-driven setting. Another recent work (Cheung et al., 2020) studies the non-stationary problem and proposes dual-based algorithms that utilize the trajectories sampled from the prior estimate distribution, which can be classified as a static policy (see Section 4.3). We show a major disadvantage of the static policy is that when directly applied to the data-driven setting with estimation errors, it could incur linear regret even when W_T is sublinear; and also the Wasserstein distance is in general tighter than the total variation distance used therein.

For the uninformative setting, we assume no prior knowledge on the true distribution. This setting is consistent with the unknown distribution setting in the literature of OLP problem (Molinaro and Ravi, 2013; Agrawal et al., 2014; Gupta and Molinaro, 2014) and the setting of blind NRM (Besbes and Zeevi, 2012; Jasin, 2015). We modify the WBDB by replacing the prior estimate of each distribution with their uniform mixture distribution to propose a new measure called *Wasserstein-*

based non-stationarity measure (WBNB). By its definition, the WBNB captures the cumulative deviation for all the distributions from their centric distribution, and it thus reflects the intensity of the non-stationarity of the distributions. In this sense, the WBNB concerns the global change of the distributions, whereas the previous non-stationarity measures (Besbes et al., 2014, 2015; Cheung et al., 2019) in an unconstrained setting characterize the local and temporal change of the distributions over time. Such temporal change measures actually fail in a constrained setting. Thus it addresses the necessity of such a global measure and reveals the interaction between the constraints and the non-stationary environment. Non-stationarity measure is further considered in Liu et al. (2022).

As a probability distance metric, the Wasserstein distance has been widely used as a measure of the deviation between estimate and true distribution in the distributionally robust optimization literature (e.g. Esfahani and Kuhn (2018)) to represent confidence set and it has demonstrated good performance both theoretically and empirically. To the best of our knowledge, we are the first to use the Wasserstein distance in an online optimization context (Jiang et al., 2020). From a modeling perspective, the two proposed measures WBDB and WBNB contribute to the study of non-stationary environment for online optimization problem. Specifically, the data-driven setting relaxes the common assumption adopted in the NRM literature that the true distributions are known to the decision maker by allowing the prior estimates to deviate from the true distributions. This deviation can be interpreted as an estimation or model misspecification error, and WBDB establishes a connection between the deviation and algorithmic performance. The uninformative setting generalizes a stream of online learning literature (e.g. Besbes et al. (2015)), which mainly concerned with the unconstrained settings and includes bandits problem (Garivier and Moulines, 2008; Besbes et al., 2014) and reinforcement learning problem (Cheung et al., 2019; Lecarpentier and Rachelson, 2019) as special cases. WBDB adds to the current dictionary of non-stationarity definitions and it specializes for a characterization of the constrained setting.

New Result II: Primal-Dual Algorithm for the Allocation Problem with Service Level

Constraints. We consider a resource allocation problem with service level constraints. Motivated by the primal-dual online algorithm, we propose a simple rationing policy, called the Max-Weighted-Service policy. Our policy assigns a random weight to each customer and based on the weights we solve, after demand realization, a deterministic capacity allocation and demand fulfillment problem to maximize a weighted service measure function. The random weight is sampled from a sufficiently large set that can be constructed offline.

Our main result is to show the Max-Weighted-Service policy is asymptotically optimal for a very general class of capacity allocation and demand fulfillment problems with individual service constraints. The generality of our model formulation is similar to the so-called newsvendor networks models proposed by Mieghem and Rudi (2002) (see also Bassamboo et al. (2010)). Indeed, our model is applicable in inventory pooling, process flexibility, assemble to order, transshipment, substitution, etc, as long as the fulfilled demand can be modeled as a linear transformation of capacity.

Our result is derived based on a new semi-infinite linear programming formulation of the problem. After proving the strong duality result of the semi-infinite linear programming, we develop a minimax stochastic programming formulation and apply an existing first-order algorithm to compute an optimal or near-optimal capacity level. The algorithm converges to a globally optimal solution whenever the objective function is convex, as predicted by existing theory. When the objective function is non-convex, we propose heuristics to compute near-optimal solutions.

Comparison with literature. Unlike the models of Mieghem and Rudi (2002) that penalize unsatisfied demand in the cost function, our model explicitly imposes individual service constraint for each customer. And the service constraints can be defined in a variety of ways. Indeed, our Max-Weighted-Service policy is asymptotically optimal under very mild conditions on the service measure functions, which are satisfied by both Type I and Type II service levels. Type I and Type II service levels have usually been analyzed separately in the literature. Our approach allows a unified treatment for both service levels, and beyond. Besides these two metrics that

are commonly used in practice and studied in the literature, we also allow the service level of a customer to depend on, for example, the probability that its demand is fully satisfied as well as the probability that a certain fraction of its demand is satisfied.

Despite the generality of the model, our approach to derive the policy is simple. We formulate the problem of finding an optimal randomized policy, for a fixed capacity level, as a semi-infinite linear program. The decision variable can be interpreted as the probability measure over the set of all possible deterministic policies, not just all possible priority lists. (Priority policies are not always optimal for our general model.) Although this formulation is natural, it appears to be new in the literature that addresses individual service constraints. Randomized policies have been studied for various special cases of our model, see for example Swaminathan and Srinivasan (1999), Alptekinoğlu et al. (2013), Zhong et al. (2017), Lyu et al. (2019), but their formulations are different than ours. For example, for inventory pooling with Type I service constraints, Swaminathan and Srinivasan (1999) partition the support of demand into different regions, and the decision variable is, for each demand region, the probability of choosing a particular priority list. Alptekinoğlu et al. (2013)) takes a similar approach. We discuss the difference between our approach and those of Zhong et al. (2017) and Lyu et al. (2019) below.

For the special cases of single-resource pooling and process flexibility without second-stage allocation costs, our policy recovers those in Hou et al. (2009), Zhong et al. (2017), and Lyu et al. (2019). However, our policy is derived using a different approach. For example, Lyu et al. (2019) uses the infinite-time horizon model to study the single-period model. They derive an allocation policy for the infinite-time horizon model. They then use their policy to derive a sufficient condition for a given capacity being feasible for the infinite-time horizon model, which corresponds to the notion of asymptotically feasible defined in our paper for the single-period model. Their analysis appears to be specific to the process flexibility problem and is much more involved than the analysis of the single-resource allocation problem in Zhong et al. (2017). In contrast, our semi-infinite linear programming formulation and its dual allow us to derive the necessary and

sufficient condition of a given capacity level for the single-period problem directly. Our allocation policy is motivated by applying the stochastic gradient descent (SGD) algorithm to the dual problem. Moreover, the SGD-based approach proves the asymptotic optimality of the policy for a much more general model that even includes second-stage allocation costs, while Blackwell's approachability theorem focuses on finding a feasible path to approach the target set without concerning optimality of this path with regard to the allocation cost. Also, the connection between the dual variable and the optimal allocation policy appears to be new.

New Result III: An Online Learning Policy for Inventory Control. We consider a stochastic lost-sales inventory control system with a lead time L over a planning horizon T (Chen et al., 2022). Supply is uncertain, and is a function of the order quantity (due to random yield/capacity, etc). We aim to minimize the T-period cost, a problem that is known to be computationally intractable even under known distributions of demand and supply. We assume that both the demand and supply distributions are unknown and develop a computationally efficient online learning algorithm to learn the optimal constant-order policy using historical data. We prove that the cost incurred by our learning algorithm is higher than that of the optimal constant-order policy by at most $O(L + \sqrt{T})$ for any $L \geqslant 0$. On the other hand, it is shown in Bu et al. (2020) that the cost of the optimal constant-order policy converges to the optimal policy exponentially fast in the lead time L, implying that the cost of our learning algorithm is higher than the *optimal policy* by at most $O(L+\sqrt{T})$ when $L \geqslant O(\log T)$. This is the first learning algorithm that provably approaches the optimal policy.

In fact, even for the special case of lost-sales inventory system with lead times and deterministic supply, which has been extensively studied in the learning literature, there is no existing learning algorithm that approaches the optimal policy under any parameter regime. Algorithms developed for this special case in Huh et al. (2009), Zhang et al. (2020), Agrawal and Jia (2022) and Lyu et al. (2021) are designed to approach various heuristic policies, and the best regret convergence rate in terms of its dependence in L and T is $O(L\sqrt{T})$, also benchmarked against certain

heuristic policy. Our regret rate of $O(L+\sqrt{T})$ dominates the rate of $O(L\sqrt{T})$, not to mention that our regret also holds with supply uncertainty, and when $L \geqslant O(\log T)$, our regret is benchmarked against the optimal policy.

Comparison to literature. The area of inventory control with online demand learning has been flourishing in recent years (Lim et al., 2006). However, all existing studies assume supply is deterministic and demand is the only source of uncertainties. Algorithms are then proposed to learn only the demand distribution, which cannot be directly applied to the case when supply also has uncertainties. A special case of our problem is the well-known lost-sales inventory system with positive lead times L, but with deterministic supply. Because this special case is already too complex to have tractable optimal solutions, researchers propose online learning algorithms to learn various heuristics. Huh et al. (2009) propose a gradient based learning algorithm that converges to the optimal base-stock heuristic policy. Their results are then improved by the SGD based learning algorithm in Zhang et al. (2020) whose cost is proved to be higher than the optimal base-stock heuristic policy by at most $O(\exp(L)\sqrt{T})$. Agrawal and Jia (2022) develops a bisection based algorithm and further improves the result to $O(L\sqrt{T})$. By developing a UCB based learning algorithm for discrete demand, Lyu et al. (2021) proves that the cost of their algorithm is higher than the optimal capped base-stock heuristic policy by at most $O(L\sqrt{T})$. Different from the existing results, when $L \geqslant O(\log T)$, the regret of our algorithm, $O(L+\sqrt{T})$, is obtained by comparing to the *true optimal policy*. Other inventory control models with online demand learning include Lin et al. (2022) considering the repeated newsvendor problem and characterizing how the regret depends on a separation assumption about the demand, Huh and Rusmevichientong (2009) considering the lost-sales inventory system with zero lead time, Zhang et al. (2018) considering perishable products, Chen and Shi (2019) studying the dual sourcing inventory system, Yuan et al. (2021) exploring inventory control problems with fixed setup cost. These works all develop SGD based learning algorithms and achieve a regret rate of $O(\sqrt{T})$. Problems with joint pricing and inventory decisions are explored in Chen et al. (2019) and Chen et al. (2021) with regret $O(\sqrt{T})$.

Part I

Online Algorithms with Competitive Analysis

2 | Tight LP-based Guarantees

In this chapter, we present details for the online algorithms with tight LP-based guarantees for online stochastic knapsack and k-unit prophet inequalities, which are fundamental problems in online algorithms, often applied as crucial subroutines in mechanism design and general online resource allocation. Section 2.1 formalizes the model/notation for the general problem. Section 2.2 comes up with the general LP framework to work with. Section 2.3 presents all of our results for k-unit prophet inequalities. Section 2.4 considers the single-resource problem under the knapsack setting. Section 2.5 presents our improvement in the unit-density special case. All the formal proofs are deferred to the appendix, however, the proof techniques are always sketched.

2.1 Problem Formulation

We formalize the model and the notations. We consider an online resource allocation problem over T discrete time periods. We assume that there is a single resource where the initial capacity is without loss of generality scaled to 1. At the beginning of each period $t = 1, 2, \ldots, T$, a query arrives, denoted by query t, and is associated with a non-negative stochastic reward $\tilde{r}_t$ and a positive stochastic size $\tilde{d}_t$, where $(\tilde{r}_t, \tilde{d}_t)$ is assumed to follow a joint distribution $F_t(\cdot)$ that can be inhomogeneous in time and is independent across time. After the value of $(\tilde{r}_t, \tilde{d}_t)$ is revealed, the decision maker has to decide irrevocably whether to serve or reject this query. If served, query t will take up $\tilde{d}_t$ capacity of the resource and a reward $\tilde{r}_t$ will be collected. If rejected, then

no reward is collected and no resource is consumed. The goal is to maximize the total collected reward subject to the capacity constraint of the resource.

We classify the elements described above into two categories: (i) the problem instance $I = (F_1, F_2, \ldots, F_T)$ denoting all distributions; and (ii) the realization of the reward and size of the arriving queries $H = ((\tilde{r}_1, \tilde{d}_1), \ldots, (\tilde{r}_T, \tilde{d}_T))$.

Any online policy π for the decision maker can be specified by a set of decision variables $\{x_t^\pi\}_{t=1,\ldots,T}$, where x_t^π is a binary variable and denotes whether query t is served by the resource. Note that if π is a randomized policy, then x_t^π is a random binary variable. A policy π is *feasible* if and only if π is non-anticipative; i.e., for each t, the value of x_t^π can only depend on the problem instance I and the rewards and the sizes of the arriving queries up to query t, denoted by $\{(\tilde{r}_1, \tilde{d}_1), \ldots, (\tilde{r}_t, \tilde{d}_t)\}$. Moreover, π needs to satisfy the capacity constraint $\sum_{t=1}^T \tilde{d}_t \cdot x_t^\pi \leq 1$ and the constraint $x_t^\pi \leq 1$ for all $t = 1, \ldots, T$. The total collected reward of policy π is denoted by $V^\pi(H) = \sum_{t=1}^T \tilde{r}_t \cdot x_t^\pi$, and the expected total collected reward of policy π is denoted by $\mathbb{E}_{\pi, H\sim I}[V^\pi(H)]$.

We compare to the prophet, who can make decisions based on the knowledge of the realizations of all the queries. Similarly, the optimal offline decision is specified by $\{x_t^{\text{off}}(H)\}_{t=1,\ldots,T}$, which is the optimal solution to the following offline problem

$$\begin{aligned} V^{\text{off}}(H) = \max \quad & \sum_{t=1}^T \tilde{r}_t \cdot x_t \\ \text{s.t.} \quad & \sum_{t=1}^T \tilde{d}_t \cdot x_t \leq 1 \\ & x_t \in \{0,1\}, \quad \forall t. \end{aligned} \tag{2.1}$$

The prophet's value is denoted by $\mathsf{Proph}(I) = \mathbb{E}_{H\sim I}[V^{\text{off}}(H)]$. For any feasible online policy π,

its competitive ratio γ is defined as

$$\gamma = \inf_{I} \frac{\mathbb{E}_{\pi, H\sim I}[V^{\pi}(H)]}{\mathsf{Proph}(I)}. \tag{2.2}$$

In this paper, however, instead of directly comparing to $\mathsf{Proph}(I)$, we compare to a linear programming (LP) upper bound of $\mathsf{Proph}(I)$, which will be useful for decomposing into single-resource subproblems. We consider the following LP as an upper bound:

$$\begin{aligned} \mathsf{ExAnte}(I) = \max \quad & \sum_{t=1}^{T} \mathbb{E}_{(\tilde{r}_t, \tilde{d}_t)\sim F_t}[\tilde{r}_t \cdot x_t(\tilde{r}_t, \tilde{d}_t)] \\ \text{s.t.} \quad & \sum_{t=1}^{T} \mathbb{E}_{(\tilde{r}_t, \tilde{d}_t)\sim F_t}[\tilde{d}_t \cdot x_t(\tilde{r}_t, \tilde{d}_t)] \le 1 \\ & x_t(r_t, d_t) \ge 0, \quad \forall t, \forall (r_t, d_t). \end{aligned} \tag{2.3}$$

Here $I = (F_1, \ldots, F_T)$ denotes the problem setup, while $x_t(r_t, d_t)$ denotes the probability of serving query t conditional on its reward and size realizing to r_t and d_t, respectively. The following lemma shows that for a policy π to have a competitive ratio of at least γ, it suffices that $\mathbb{E}_{\pi, H\sim I}[V^{\pi}(H)]$ is at least γ times the optimal fractional allocation value $\mathsf{ExAnte}(I)$ for every problem instance I.

Lemma 2.0.1. *For all problem instances I, it holds that* $\mathsf{Proph}(I) \le \mathsf{ExAnte}(I)$.

Remark. By considering a LP upper bound (2.3), we can easily generalize any competitive ratio for our problem to a general multi-resource allocation problem, where there are multiple resources each with an initial capacity, and the query will be assigned to be served by one resource, with resource-dependent reward and size. This LP-based generalization approach has been applied in the previous literature (Alaei et al., 2012, 2013; Wang et al., 2018; Stein et al., 2020) and we refer interested readers to Jiang et al. (2022a) for details.

2.2 LP DUALITY.

We now show that in order to obtain the tight guarantee, it is sufficient to work with a new LP, as well as its dual. We begin with considering the optimal online policy given by dynamic programming (DP). Denote by $V_t^*(c_t)$ the "value-to-go" function at period t with remaining capacity c_t. We have the following backward induction:

$$V_t^*(c_t) = V_{t+1}^*(c_t) + \mathbb{E}_{(\tilde{r}_t,\tilde{d}_t)}[\underbrace{\max\left\{0, 1_{\{c_t \geq \tilde{d}_t\}} \cdot (\tilde{r}_t + V_{t+1}^*(c_t - \tilde{d}_t) - V_{t+1}^*(c_t))\right\}}_{\text{marginal increase}}], \ \forall t, 0 \leq c_t \leq 1$$

where $V_{T+1}^*(\cdot) = 0$. Having this notation, we are interested in characterizing the quantity $\inf_I \frac{V_1^*(1)}{\mathsf{ExAnte}(I)}$. Note that under the constraint $\sum_{t=1}^T \mathbb{E}[\tilde{d}_t] \leq 1$, we must have $\mathsf{ExAnte}(I) = \sum_{t=1}^T \mathbb{E}_{(\tilde{r}_t,\tilde{d}_t)}[\tilde{r}_t]$. We scale the values of $\tilde{r}_t$ so that $\mathsf{ExAnte}(I) = \sum_{t=1}^T \mathbb{E}_{(\tilde{r}_t,\tilde{d}_t)}[\tilde{r}_t] = 1$. We denote by $p(r_t, d_t) = P((\tilde{r}_t, \tilde{d}_t) = (r_t, d_t))$ and $p_t(d_t) = \sum_{r_t} p(r_t, d_t)$. With some further massaging, we can re-express $\inf_I \frac{V_1^*(1)}{\mathsf{ExAnte}(I)}$ as the infimum value of a LP Primal$(\boldsymbol{p}, \boldsymbol{D})$, as well as its dual which we denote by Dual$(\boldsymbol{p}, \boldsymbol{D})$, over $\boldsymbol{p} = \{p_t(d_t)\}_{\forall t, \forall d_t}$ and $\boldsymbol{D} = \{d_t\}_{\forall t, \forall d_t}$ such that $\sum_{t=1}^T \sum_{d_t} p_t(d_t) \cdot d_t \leq 1$. The following lemma formalizes this fact.

Lemma 2.0.2. *It holds that*

$$\inf_I \frac{V_1^*(1)}{\mathsf{ExAnte}(I)} = \inf_{(\boldsymbol{p},\boldsymbol{D}):\sum_{t=1}^T \sum_{d_t} p_t(d_t)\cdot d_t \leq 1} \mathit{Dual}(\boldsymbol{p}, \boldsymbol{D}).$$

where the formulations of Dual$(\boldsymbol{p}, \boldsymbol{D})$ are given in (2.4).

The formulation of Dual($\boldsymbol{p}, \boldsymbol{D}$) is given as follows.

$$\text{Dual}(\boldsymbol{p}, \boldsymbol{D}) = \max\ \theta \tag{2.4}$$

$$\text{s.t. } \theta \cdot p_t(d_t) \le \sum_{c_t : c_t \ge d_t} \alpha_t(d_t, c_t), \quad \forall t, \forall d_t \tag{2.4a}$$

$$\alpha_t(d_t, c_t) \le p_t(d_t) \cdot \sum_{\tau \le t-1} \sum_{d_\tau} (\alpha_\tau(d_\tau, c_t + d_\tau) - \alpha_\tau(d_\tau, c_t)), \ \forall t, \forall d_t, \forall c_t < 1 \tag{2.4b}$$

$$\alpha_t(d_t, 1) \le p_t(d_t) \cdot (1 - \sum_{\tau \le t-1} \sum_{d_\tau} \alpha_\tau(d_\tau, 1)), \quad \forall t, \forall d_t \tag{2.4c}$$

$$\theta, \alpha_t(d_t, c_t) \ge 0, \quad \forall t, \forall d_t, \forall c_t, \quad \alpha_t(d_t, c_t) = 0, \quad \forall t, \forall d_t, \forall c_t > 1$$

Equivalence to the Online Contention Resolution Scheme (OCRS) problems. In fact, for general $\boldsymbol{p}$ and $\boldsymbol{D}$, Dual($\boldsymbol{p}, \boldsymbol{D}$) characterizes the knapsack OCRS problem described as follows.

Definition 2.1 (Knapsack OCRS Problem)**.** There is a sequence of queries $t = 1, \ldots, T$, and for each query t, its size is independently randomly drawn from a known subset of [0,1]. For each possible size $d \in [0, 1]$ we let $p_t(d)$ denote the probability of query t having size d. After the query's size is observed, the query must be immediately served or rejected. The total size of queries served cannot exceed 1, and it is promised that the total expected size is at most 1, i.e. $\sum_t \sum_d p_t(d) \cdot d \leqslant 1$. The goal of an online algorithm is to serve every query t with probability at least γ conditional on the size being realized to d, for each possible d, and for a constant $\gamma \in [0, 1]$ as large as possible.

To see this, we interpret the variable θ as the guarantee γ in the knapsack OCRS problem, and we interpret the variable $\alpha_t(d_t, c_t)$ for each t, each d_t, and each $0 \le c_t \le 1$ as the *ex-ante* probability (the unconditional probability) that query t becomes active with size d_t and is served when the remaining capacity is c_t at the beginning of period t. Then, constraint (2.4a) corresponds to each query t being served with probability at least $\theta = \gamma$ conditional on it being active with

size d_t. Moreover, for each t and each $0 \leq c_t < 1$, the term $\sum_{\tau=1}^{t-1} \sum_{d_\tau} \alpha_\tau(d_\tau, c_t + d_\tau)$ denotes the ex-ante probability that the remaining capacity has "reached" the state c_t in the first $t-1$ periods and the term $\sum_{\tau=1}^{t-1} \sum_{d_\tau} \alpha_\tau(d_\tau, c_t)$ denotes the ex-ante probability that the remaining capacity has "left" the state c_t. Thus, the difference $\sum_{\tau=1}^{t-1} \sum_{d_\tau} (\alpha_\tau(d_\tau, c_t + d_\tau) - \alpha_\tau(d_\tau, c_t))$ denotes exactly the ex-ante probability that the remaining capacity at the beginning of period t is c_t. Similarly, the difference $1 - \sum_{\tau=1}^{t-1} \sum_{d_\tau} \alpha_\tau(d_\tau, 1)$ denotes the ex-ante probability that the remaining capacity at the beginning of period t is 1. Since each query t can only be served after being active with size d_t, which happens independently with probability $p_t(d_t)$, we recover constraint (2.4b) and constraint (2.4c).

When each d_t in $\boldsymbol{D}$ only takes value 0 or $1/k$, Dual($\boldsymbol{p}, \boldsymbol{D}$) characterizes the k-unit OCRS problem described as follows.

Definition 2.2 (k-unit OCRS Problem)**.** There is a sequence of queries $t = 1, \ldots, T$, each of which is active independently according to a known probability p_t. Whether a query is active is sequentially observed, and active queries can be immediately served or rejected, while inactive queries must be rejected. At most k queries can be served in total, and it is promised that $\sum_t p_t \leqslant k$. The goal of an online algorithm is to serve every query t with probability at least γ conditional on it being active, for a constant $\gamma \in [0, 1]$ as large as possible, potentially with the aid of randomization.

Following the discussions above, the duality between Dual($\boldsymbol{p}, \boldsymbol{D}$) and Primal($\boldsymbol{p}, \boldsymbol{D}$) implies that OCRS is the dual problem of our online resource allocation problem, under both the k-unit setting (in which case our online resource allocation problem becomes the k-unit prophet inequality problem) and the knapsack setting. Consequently, the worst-case guarantee for the OCRS problem is equivalent to the worst-case guarantee of the optimal online policy relative to the LP relaxation.

2.3 Multi-unit Prophet Inequalities

We now consider the multi-unit setting. To be specific, we assume that $d_t = \frac{1}{k}$ for an integer k, for each t, i.e., the online decision maker can serve at most k queries to collect the corresponding rewards. Note that when $k = 1$, our problem reduces to the well-known prophet inequality (Krengel and Sucheston, 1978), and for general k, we get the so-called multi-unit prophet inequalities (Alaei, 2011). The main result of this section is that for each k, we derive the tight competitive ratio for the k-unit prophet inequality problem with respect to the LP upper bound, or equivalently the optimal solution γ_k^* to the k-unit OCRS problem. Note that our values γ_k^* strictly exceed $1 - \frac{1}{\sqrt{k+3}}$ for all $k > 1$, and hence we also improve the best-known prophet inequalities for all $k > 1$. The structure of our proof follows the three steps outlined in Section 1.1.

2.3.1. LP formulation of k-unit OCRS problem. We first present a new LP formulation of the k-unit OCRS problem, assuming that the vector $\boldsymbol{p}$ is given in advance. This LP is derived from taking the dual of the LP formulation of the optimal dynamic programming (DP) policy. Thus, we name our LP as Dual$(\boldsymbol{p}, k)$, which is exactly Dual$(\boldsymbol{p}, \boldsymbol{D})$ in (2.4) under the k-unit setting. Then, the restriction $\sum_{t=1}^{T} \sum_{d_t} p_t(d_t) \cdot d_t \leq 1$, which follows from **??**, can be translated into $\sum_{t=1}^{T} p_t \leq k$ by noting d_t takes a single value $1/k$ when query t becomes active.

$$
\begin{aligned}
\text{Dual}(\boldsymbol{p}, k) = \max \quad & \theta && (2.5)\\
\text{s.t.} \quad & \theta \leq \frac{\sum_{l=1}^{k} x_{l,t}}{p_t} \quad \forall t && (2.5a)\\
& x_{1,t} \leq p_t \cdot (1 - \sum_{\tau<t} x_{1,\tau}) \quad \forall t && (2.5b)\\
& x_{l,t} \leq p_t \cdot \sum_{\tau<t} (x_{l-1,\tau} - x_{l,\tau}) \quad \forall t, \forall l = 2, \ldots, k && (2.5c)\\
& x_{1,t} \geq 0, x_{2,t} \geq 0, \ldots, x_{k,t} \geq 0.
\end{aligned}
$$

Here, the variable θ can be interpreted as guarantee γ in the k-unit OCRS problem and $x_{l,t}$ can be interpreted as the ex-ante probability of serving query t as the l-th one. Then, constraint (2.5a) guarantees that each query t is served with an ex-ante probability $\theta \cdot p_t$. Moreover, it is easy to see that the term $\sum_{\tau<t} x_{l-1,\tau}$ can be interpreted as the probability that the number of served queries has "reached" $l-1$ during the first $t-1$ periods, while the term $\sum_{\tau<t} x_{l,\tau}$ can be interpreted as the probability that the number of served queries is larger than $l-1$. Then, the term $\sum_{\tau<t}(x_{l-1,\tau}-x_{l,\tau})$ denotes the probability that the number of served queries is $l-1$ at the beginning of period t. Similarly, the term $1-\sum_{\tau<t} x_{1,\tau}$ denotes the probability that no query is served at the beginning of period t. Further note that each query t can be served only after it becomes active, which happens independently with probability p_t, and hence we get constraint (2.5b) and (2.5c).

Although presented in a different context, the "γ-Conservative Magician" procedure of Alaei (2011) implies a feasible solution to Dual$(\boldsymbol{p}, k)$, for any $\boldsymbol{p}$ and any k satisfying $\sum_{t=1}^{T} p_t \le k$. We now describe this implied solution in Definition 2.3, which is based on a predetermined θ.

Definition 2.3. A candidate solution to Dual$(\boldsymbol{p}, k)$ in (2.5) given θ (potentially infeasible)

1. For a fixed $\theta \in [0,1]$, we define $x_{1,t}(\theta) = \theta \cdot p_t$ from $t = 1$ up to $t = t_2$, where t_2 is defined as the first time among $\{1,\dots,T\}$ such that $\theta > 1 - \sum_{t=1}^{t_2} \theta \cdot p_t$ and if such a t_2 does not exist, we denote $t_2 = T$. Then we define $x_{1,t}(\theta) = p_t \cdot (1 - \sum_{\tau=1}^{t-1} x_{1,\tau}(\theta))$ from $t = t_2+1$ up to $t = T$.

2. For $l = 2, 3, \dots, k-1$, we do the following:

   (a) Define $x_{l,t}(\theta) = 0$ from $t = 1$ up to $t = t_l$.

   (b) Define $x_{l,t}(\theta) = \theta \cdot p_t - \sum_{v=1}^{l-1} x_{v,t}(\theta)$ from $t = t_l + 1$ up to $t = t_{l+1}$, where t_{l+1} is defined as the first time among $\{1,\dots,T\}$ such that

$$\theta \cdot p_{t_{l+1}+1} - \sum_{v=1}^{l-1} x_{v,t_{l+1}+1}(\theta) > p_{t_{l+1}+1} \cdot \sum_{t=1}^{t_{l+1}} (x_{l-1,t}(\theta) - x_{l,t}(\theta))$$

   and if such a t_{l+1} does not exist, we denote $t_{l+1} = T$.

(c) Define $x_{l,t}(\theta) = p_t \cdot \sum_{\tau=1}^{t-1}(x_{l-1,\tau}(\theta) - x_{l,\tau}(\theta))$ from $t = t_{l+1} + 1$ up to $t = T$.

3. Define $x_{k,t}(\theta) = 0$ for $t = 1, \ldots, t_k$ and define $x_{k,t}(\theta) = \theta \cdot p_t - \sum_{v=1}^{k-1} x_{v,t}(\theta)$ for $t = t_k+1, \ldots, T$.

Note that in the above construction, the values of $\{t_l\}$ and $\{x_{l,t}(\theta)\}$ are uniquely determined by θ. Obviously, for an arbitrary θ, the solution $\{\theta, x_{l,t}(\theta)\}$ is not necessarily a feasible solution to Dual($\boldsymbol{p}, k$), much less an optimal one. Alaei et al. (2012) shows that if we set $\theta = 1 - \frac{1}{\sqrt{k+3}}$, then $\{\theta, x_{l,t}(\theta)\}$ is a feasible solution to Dual($\boldsymbol{p}, k$). Thus, they obtain a $1 - \frac{1}{\sqrt{k+3}}$-lower bound of the optimal competitive ratio. However, we now try to identify a θ^*, which is *dependent on* $\boldsymbol{p}$, such that $\{\theta^*, x_{l,t}(\theta^*)\}$ is an *optimal* solution to Dual($\boldsymbol{p}, k$).

2.3.2. Characterizing the optimal LP solution for a given $\boldsymbol{p}$. We begin by proving the condition on θ for $\{\theta, x_{l,t}(\theta)\}$ to be a feasible solution to Dual($\boldsymbol{p}, k$).

Lemma 2.3.1. *For any vector $\boldsymbol{p}$, there exists a unique $\theta^* \in [0, 1]$ such that $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta^*) = 1 - \theta^*$. Moreover, for any $\theta \in [0, \theta^*]$, $\{\theta, x_{l,t}(\theta)\}$ is a feasible solution to Dual($\boldsymbol{p}, k$).*

We now prove that $\{\theta^*, x_{l,t}(\theta^*)\}$ is an optimal solution to Dual($\boldsymbol{p}, k$). The dual of Dual($\boldsymbol{p}, k$) is denoted by Primal($\boldsymbol{p}, k$). To prove the optimality of $\{\theta^*, x_{l,t}(\theta^*)\}$, we will construct a feasible dual solution $\{\beta^*_{l,t}, \xi^*_t\}$ to Primal($\boldsymbol{p}, k$) such that complementary slackness conditions hold for the primal-dual pair $\{\theta^*, x_{l,t}(\theta^*)\}$ and $\{\beta^*_{l,t}, \xi^*_t\}$; then, the well-known primal-dual optimality criterion (Dantzig and Thapa, 2006) establishes that $\{\theta^*, x_{l,t}(\theta^*)\}$ and $\{\beta^*_{l,t}, \xi^*_t\}$ are the optimal primal-dual pair to Dual($\boldsymbol{p}, k$) and Primal($\boldsymbol{p}, k$), which completes our proof. The above arguments are formalized in the following theorem.

Theorem 2.4. *The solution $\{\theta^*, x_{l,t}(\theta^*)\}$ is optimal for Dual($\boldsymbol{p}, k$), where θ^* is the unique solution to $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta^*) = 1 - \theta^*$.*

Theorem 2.4 shows that Theorem 2.3 constructs an optimal solution to Dual($\boldsymbol{p}, k$), *as long as the θ is set as the optimal θ^**, as defined in Lemma 2.3.1. This optimal θ^* is uniquely defined based on $\boldsymbol{p}$. Lemma 2.3.1 further shows that any $\theta \leqslant \theta^*$ is feasible, and hence if we can find a θ that

is no greater than the θ^* arising from any $\boldsymbol{p}$, then Theorem 2.3 will correspond to a $\boldsymbol{p}$-agnostic procedure for the k-unit prophet inequality or OCRS problem with a guarantee of θ.

2.3.3. Characterizing the worst-case distribution. Our goal is now to find the $\boldsymbol{p}$ such that the optimal objective value θ^* of Dual$(\boldsymbol{p}, k)$ in (2.5) reaches its minimum. We would like to characterize the worst-case distribution and then compute the competitive ratio.

We first characterize the worst-case distribution for which the optimal objective value of Dual$(\boldsymbol{p}, k)$ reaches its minimum. Obviously, it is enough for us to consider only the $\boldsymbol{p}$ satisfying $\sum_{t=1}^{T} p_t = k$. We show in the following lemma that splitting one query into two queries can only make the optimal objective value of Dual$(\boldsymbol{p}, k)$ become smaller, and thus, in the worst-case distribution, each p_t should be infinitesimally small.

Lemma 2.4.1. *For any $\boldsymbol{p} = (p_1, \ldots, p_T)$ satisfying $\sum_{t=1}^{T} p_t = k$, and any $\sigma \in [0, 1]$, $1 \le q \le T$, if we define a new sequence of arrival probabilities $\tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_{T+1})$ such that*

$$\tilde{p}_t = p_t \quad \forall t < q, \quad \tilde{p}_q = p_q \cdot \sigma, \quad \tilde{p}_{q+1} = p_q \cdot (1 - \sigma) \ \text{ and } \ \tilde{p}_{t+1} = p_t \quad \forall q + 1 \le t \le T,$$

then it holds that Dual$(\boldsymbol{p}, k) \ge$ Dual$(\tilde{\boldsymbol{p}}, k)$.

Now, for each $\boldsymbol{p} = (p_1, \ldots, p_T)$ satisfying $\sum_{t=1}^{T} p_t = k$, we assume without loss of generality that p_t is a rational number for each t, i.e., $p_t = \frac{n_t}{N}$ where n_t is an integer for each t and N is an integer denoting the common denominator. We first split p_1 into $\frac{1}{N}$ and $\frac{n_1-1}{N}$ to form a new sequence of arrival probabilities. By Lemma 2.4.1, we know that such an operation can only decrease the optimal objective value of Dual$(\boldsymbol{p}, k)$. We then split $\frac{n_1-1}{N}$ into $\frac{1}{N}$ and $\frac{n_1-2}{N}$ and so on. In this way, we split p_1 into n_1 copies of $\frac{1}{N}$ to form a new sequence of arrival probabilities and Lemma 2.4.1 guarantees that the optimal objective value of Dual$(\boldsymbol{p}, k)$ can only become smaller. We repeat the above operation for each t. Finally, we form a new sequence of arrival probabilities, denoted by $\boldsymbol{p}^N = (\frac{1}{N}, \ldots, \frac{1}{N}) \in \mathbb{R}^{Nk}$, and we have Dual$(\boldsymbol{p}, k) \ge$ Dual$(\boldsymbol{p}^N, k)$. Intuitively, when $N \to \infty$, then the optimal objective value of Dual$(\boldsymbol{p}, k)$ reaches its minimum. Note that when

$N \to \infty$, we always have $\sum_{t=1}^{Nk} p_t^N = k$, and then the Bernoulli arrival process approximates a Poisson process with rate 1 over the time interval $[0, k]$. The above argument implies that the worst-case arrival process is a Poisson process.

Under the Poisson process, for each fixed ratio $\theta \in [0, 1]$, our solution in Definition 2.3 can be interpreted as a solution to an ordinary differential equation (ODE). We further note that for $\boldsymbol{p}^N$ and any $\theta \in [0, 1]$, our solution in Definition 2.3 can be regarded as the solution obtained from applying Euler's method to solve this ODE by uniformly discretizing the interval $[0, k]$ into Nk discrete points. Then, for any fixed ratio $\theta \in [0, 1]$, after showing the Lipschitz continuity of the function defining this ODE, we can apply the global truncation error theorem of Euler's method (Theorem 212A in Butcher and Goodwin (2008)) to establish the solution under the Poisson process as the limit of the solution under $\boldsymbol{p}^N$ when $N \to \infty$. Based on this convergence, we can prove that the optimal value under the Poisson process is equivalent to $\lim_{N\to\infty} \text{Dual}(\boldsymbol{p}^N, k)$, which is the optimal ratio we are looking for.

Specifically in the case of $k = 2$, we construct an example showing that relative to the weaker prophet benchmark $\text{Proph}(I)$, it is not possible to do much better than γ_2^*. Our construction is based on adapting the tight example relative to the stronger benchmark $\text{ExAnte}(I)$). We note that this suggests that there is *some separation* between optimal ex-ante vs. non-ex-ante prophet inequalities when $k > 1$, which is not the case when $k = 1$ (because they are both 1/2).

Proposition 2.5. *For the* 2*-unit prophet inequality problem,* $\sup_\pi \inf_I \frac{\mathbb{E}_{\pi, H\sim I}[V^\pi(H)]}{\text{Proph}(I)} \leq 0.6269$.

We now discuss how the construction in Definition 2.3 should be interpreted when the arrival process is a Poisson process. We find it is more convenient to work with the functions $\{\tilde{y}_{l,\theta}(\cdot)\}_{\forall l=1,\ldots,k}$ over $[0, k]$, where $\tilde{y}_{l,\theta}(t)$ denotes the ex-ante probability that there is a query served as the l-th one during the period $[0, t]$. Note that the variable $x_{l,t}(\theta)$ denotes the ex-ante probability that there is a query accepted as the l-th query at time t, and hence we have $x_{l,t}(\theta) = d\tilde{y}_{l,\theta}(t)$. We denote $\tilde{y}_{0,\theta}(t) = 1$ for each $t \in [0, k]$. Then the functions $\{\tilde{y}_{l,\theta}(\cdot)\}_{\forall l=1,\ldots,k}$

corresponding to the construction in Definition 2.3 under Poisson arrivals can be interpreted as follows.

Definition 2.6. Ordinary Differential Equation (ODE) formula under Poisson arrival

1. For each fixed $\theta \in [0,1]$, we define $\tilde{y}_{0,\theta}(t) = 1$ for each $t \in [0,k]$ and $t_1 = 0$.

2. For each $l = 1, 2, \ldots, k-1$, we do the following:

   (a) $\tilde{y}_{l,\theta}(t) = 0$ when $t \le t_l$.

   (b) When $t_l \le t \le t_{l+1}$, it holds that $\frac{d\tilde{y}_{l,\theta}(t)}{dt} = \theta - \sum_{v=1}^{l-1} \frac{d\tilde{y}_{v,\theta}(t)}{dt} = \theta - 1 + \tilde{y}_{l-1,\theta}(t), \quad \forall t_l \le t \le t_{l+1}$, where t_{l+1} is defined as the first time that $\tilde{y}_{l,\theta}(t_{l+1}) = 1-\theta$. If such a t_{l+1} does not exist, we denote $t_{l+1} = k$.

   (c) When $t_{l+1} \le t \le k$, it holds that $\frac{d\tilde{y}_{l,\theta}(t)}{dt} = \tilde{y}_{l-1,\theta}(t) - \tilde{y}_{l,\theta}(t), \quad \forall t_{l+1} \le t \le k$.

3. $\tilde{y}_{k,\theta}(t) = 0$ if $t \le t_k$ and $\frac{d\tilde{y}_{k,\theta}(t)}{dt} = \theta - 1 + \tilde{y}_{k-1,\theta}(t)$ if $t_k \le t \le k$.

Thus, by Theorem 2.4, the solution to the equation $\tilde{y}_{k,\theta}(k) = 1-\theta$ should be the minimum of the optimal objective value of Dual$(\boldsymbol{p}, k)$ in (2.5), which is the competitive ratio γ_k^* we are looking for. The above arguments are formalized in the following theorem. Note that the following Theorem 2.7 is our ultimate result for the k-unit case, while Theorem 2.6 characterizes the ODE formula mentioned in Section 1.1. We describe the computational procedure for γ_k^*.

Theorem 2.7. *For each $\theta \in [0,1]$, denote by $\{\tilde{y}_{l,\theta}(\cdot)\}$ the functions defined in Definition 2.6. Then there exists a unique $\gamma_k^* \in [0,1]$ such that $\tilde{y}_{k,\gamma_k^*}(k) = 1-\gamma_k^*$ and it holds that $\gamma_k^* = \inf_{\boldsymbol{p}}$ Dual$(\boldsymbol{p}, k)$ s.t. $\sum_{t=1}^{T} p_t = k$.*

We now show that the ODE in Definition 2.6 admits an analytical solution that enables us to compute γ_k^* for each k. For each fixed θ, when $l = 1$, it is immediate that

$$\tilde{y}_{1,\theta}(t) = \theta \cdot t, \quad \text{when } t \le t_2 = \frac{1-\theta}{\theta}, \quad \text{and} \quad \tilde{y}_{1,\theta}(t) = 1 - \theta \cdot \exp(t_2 - t), \quad \text{for } t_2 \le t \le k.$$

Now suppose that there exists a fixed $2 \le l \le k$ such that for each $1 \le v \le l-1$, it holds that

$$\tilde{y}_{v,\theta}(t) = \zeta_v + \theta \cdot t + \sum_{q=0}^{v-2} \zeta_{v,q} \cdot t^q \cdot \exp(-t), \qquad \text{when } t_v \le t \le t_{v+1}$$
$$\tilde{y}_{v,\theta}(t) = 1 + \sum_{q=0}^{v-1} \psi_{v,q} \cdot t^q \cdot \exp(-t), \qquad \text{when } t_{v+1} \le t \le k$$

for some parameters $\{\zeta_v, \zeta_{v,q}, \psi_{v,q}\}$, which are specified by θ. Then by ODE (b) and (c), it must hold that

$$\tilde{y}_{l,\theta}(t) = \zeta_l + \theta \cdot t + \sum_{q=0}^{l-2} \zeta_{l,q} \cdot t^q \cdot \exp(-t), \qquad \text{when } t_l \le t \le t_{l+1}$$
$$\tilde{y}_{l,\theta}(t) = 1 + \sum_{q=0}^{l-1} \psi_{l,q} \cdot t^q \cdot \exp(-t), \qquad \text{when } t_{l+1} \le t \le k.$$

The parameters $\{\zeta_l, \zeta_{l,q}, \psi_{l,q}\}$ can be computed in the following steps:

1. Set $\zeta_{l,l-1} = 0$ and compute $\zeta_{l,q}$ iteratively from $q = l-2$ up to $q = 0$ by setting $\zeta_{l,q} = (q+1) \cdot \zeta_{l,q+1} - \psi_{l-1,q}$.

2. Set the value of ζ_l such that $\tilde{y}_{l,\theta}(t_l) = 0$. If $l = k$, we set $t_{l+1} = k$; otherwise, we set t_{l+1} to be the solution to the following equation: $1 - \theta = \tilde{y}_{l,\theta}(t) = \zeta_l + \theta \cdot t + \sum_{q=0}^{l-2} \zeta_{l,q} \cdot t^q \cdot \exp(-t)$. Note that by definition $\tilde{y}_{l,\theta}(t)$ is monotone increasing with t, and hence we can do a bisection search on the interval $[t_l, k]$ to obtain the value of t_{l+1}.

3. Set $\psi_{l,q} = \frac{\psi_{l-1,q-1}}{q}$ for each $q = 1, \ldots, l-1$. If $l < k$, the value of $\psi_{l,0}$ is determined such that $1 - \theta = 1 + \sum_{q=0}^{l-1} \psi_{l,q} \cdot t_{l+1}^q \cdot \exp(-t_{l+1})$.

Thus, for each fixed θ, we can follow the above procedure to obtain the value of $\tilde{y}_{k,\theta}(k)$. Note that the value of $\tilde{y}_{k,\theta}(k)$ is monotone increasing with θ, and hence we can do a bisection search on $\theta \in [0, 1]$ to obtain the value of γ_k^* as the unique solution of the equation $\tilde{y}_{k,\theta}(k) = 1 - \theta$. By Theorem 2.7, γ_k^* is the optimal value for the competitive ratio.

2.4 Results for the Knapsack Setting

We present our algorithm for the general setting (knapsack problem).

2.4.1. Algorithm and interpretation. Our policy differs from existing ones for knapsack in an online setting (Dutting et al., 2020; Feldman et al., 2021; Stein et al., 2020) by eschewing the need to split queries into "large" vs. "small" based on whether its size is greater than 1/2. In fact, we can show that any algorithm which *considers large and small queries separately* in our problem is limited to $\gamma \leqslant 1/4$, and hence could not match the $\frac{1}{3+e^{-2}}$ upper bound provided earlier. The result follows by considering a problem setup I where there are 4 queries and $(\tilde{r}_t, \tilde{d}_t)$ is realized as $(\hat{r}_t, \hat{d}_t)$ with probability p_t and is realized as $(0,0)$ otherwise, for each query t, and letting $(\hat{r}_1, p_1, \hat{d}_1) = (r, 1, \epsilon)$, $(\hat{r}_2, p_2, \hat{d}_2) = (\hat{r}_3, p_3, \hat{d}_3) = (r, \frac{1-2\epsilon}{1+2\epsilon}, \frac{1}{2}+\epsilon)$, $(\hat{r}_4, p_4, \hat{d}_4) = (r/\epsilon, \epsilon, 1)$ for $r > 0$ and some small $\epsilon > 0$. We formalize the above arguments as follows.

Proposition 2.8. *If the policy π serves only either "large" queries with a size larger than 1/2, or "small" queries with a size no larger than 1/2, then it holds that* $\inf_I \frac{\mathbb{E}_{H\sim I}[V^\pi(H)]}{\mathsf{ExAnte}(I)} \leq \frac{1}{4}$.

Our policy is formally stated in Algorithm 1. Our policy is based on a pre-specified parameter γ and a salient feature of our policy is that each query t, as long as its reward and size are realized as (r_t, d_t), is included in the knapsack with a probability γ. This guarantees that our policy enjoys a competitive ratio γ. Based on γ, for each t, we use $\tilde{X}_{t-1,\gamma}$ to denote the distribution of the capacity consumption under our policy at the beginning of period t, where $\tilde{X}_{0,\gamma}$ takes value 0 deterministically. Then, for each realization (r_t, d_t) of query t, we specify a threshold $\eta_{t,\gamma}(r_t, d_t)$ such that the probability of $\tilde{X}_{t-1,\gamma} \in (\eta_{t,\gamma}(r_t, d_t), 1 - d_t]$ is smaller than or equal to γ, and the probability that $\tilde{X}_{t-1,\gamma} \in [\eta_{t,\gamma}(r_t, d_t), 1 - d_t]$ is larger than or equal to γ. When query t is realized as $(\hat{r}_t, \hat{d}_t)$, we serve query t when the realized capacity consumption is among $(\eta_{t,\gamma}(\hat{r}_t, \hat{d}_t), 1 - d_t]$, or we serve query t with a certain probability (specified in step 4), when the realized capacity consumption equals $\eta_{t,\gamma}(\hat{r}_t, \hat{d}_t)$. It is clear to see that our policy guarantees that query t is served

with a total probability γ. We finally update the distribution of capacity consumption in step 5.

Algorithm 1 Best-fit Magician policy for single-knapsack (π_γ)

1: For a fixed γ, we initialize $\tilde{X}_{0,\gamma}$ as a random variable that takes the value 0 deterministically.
2: **For** $t = 1, 2, \ldots, T$, we do the following:
3: For each realization (r_t, d_t) of $(\tilde{r}_t, \tilde{d}_t)$, we denote by $p(r_t, d_t) = P((\tilde{r}_t, \tilde{d}_t) = (r_t, d_t))$. Then, we denote a threshold $\eta_{t,\gamma}(r_t, d_t)$ satisfying:

$$P(\eta_{t,\gamma}(r_t, d_t) < \tilde{X}_{t-1,\gamma} \leq 1 - d_t) \leq \gamma \leq P(\eta_{t,\gamma}(r_t, d_t) \leq \tilde{X}_{t-1,\gamma} \leq 1 - d_t). \tag{2.6}$$

4: Denote by $(\hat{r}_t, \hat{d}_t)$ the realization of query t and by X_{t-1} the consumed capacity at the end of period $t-1$. Then we serve query t if $\eta_{t,\gamma}(\hat{r}_t, \hat{d}_t) < X_{t-1} \leq 1 - \hat{d}_t$; we serve query t with probability $\frac{\gamma - P(\eta_{t,\gamma}(\hat{r}_t, \hat{d}_t) < \tilde{X}_{t-1,\gamma} \leq 1 - \hat{d}_t)}{P(\eta_{t,\gamma}(\hat{r}_t, \hat{d}_t) = \tilde{X}_{t-1,\gamma})}$ if $X_{t-1} = \eta_{t,\gamma}(\hat{r}_t, \hat{d}_t)$.
5: Based on $\tilde{X}_{t-1,\gamma}$, for each realization (r_t, d_t) of $(\tilde{r}_t, \tilde{d}_t)$ and each point $x \in (\eta_{t,\gamma}(r_t, d_t), 1 - d_t]$, we move the $p(r_t, d_t) \cdot P(\tilde{X}_{t-1,\gamma} = x)$ measure of probability mass from point x to the new point $x + d_t$ and when $x = \eta_{t,\gamma}(r_t, d_t)$, we move the $p(r_t, d_t) \cdot (\gamma - P(\eta_{t,\gamma}(r_t, d_t) < \tilde{X}_{t-1,\gamma} \leq 1 - d_t))$ measure of probability mass from point x to the new point $x + d_t$. We obtain a new distribution and we denote by $\tilde{X}_{t,\gamma}$ a random variable with this distribution.

In fact, the policy π_γ constructs a feasible solution to Dual($\boldsymbol{p}, \boldsymbol{D}$) in (2.4). Specifically, we denote variable θ in Dual($\boldsymbol{p}, \boldsymbol{D}$) as γ in our policy π_γ. Then, for each t and each $0 \leq c_t \leq 1$, we denote the variable $\alpha_t(d_t, c_t)$ in Dual($\boldsymbol{p}, \boldsymbol{D}$) as the ex-ante probability that query t is served when the *remaining capacity* is c_t at the beginning of period t, which also denotes the amount of probability mass of $\tilde{X}_{t-1,\gamma}$ that is moved from point $1 - c_t$ to the point $1 - c_t + d_t$ when defining $\tilde{X}_{t,\gamma}$ in step 5 in Algorithm 1. Obviously, when γ is feasible, the policy π_γ guarantees that each query t with size d_t is served with an ex-ante probability $\gamma \cdot p_t(d_t)$ where $p_t(d_t) = \sum_{r_t} p(r_t, d_t)$, which corresponds to constraint (2.4a). Moreover, from step 5, it is clear to see that

$$P(\tilde{X}_{t,\gamma} = 0) = P(\tilde{X}_{t-1,\gamma} = 0) - \sum_{d_t} \alpha_t(d_t, 1)$$

and

$$P(\tilde{X}_{t,\gamma} = 1 - c) = P(\tilde{X}_{t-1,\gamma} = 1 - c) + \sum_{d_t} (\alpha_t(d_t, c + d_t) - \alpha_t(d_t, c)), \quad \forall c < 1$$

which implies that

$$P(\tilde{X}_{t-1,\gamma} = 0) = 1 - \sum_{\tau=1}^{t-1} \sum_{d_\tau} \alpha_\tau(d_\tau, 1)$$

and

$$P(\tilde{X}_{t-1,\gamma} = 1 - c) = \sum_{\tau=1}^{t-1} \sum_{d_\tau} (\alpha_\tau(d_\tau, c + d_\tau) - \alpha_\tau(d_\tau, c)), \quad \forall c < 1.$$

Then, the constraint $\alpha_t(d_t, c) \leq p_t(d_t) \cdot P(\tilde{X}_{t-1,\gamma} = 1 - c)$ for each $0 \leq c \leq 1$ in the policy π_γ corresponds to constraint (2.4b) and constraint (2.4c).

2.4.2. Proof of competitive ratio and tightness. We analyze the competitive ratio of our Best-fit Magician policy in Algorithm 1 and show that it is tight. When γ is fixed, our policy π_γ guarantees that the decision maker collects at least γ times the LP upper bound $\mathsf{ExAnte}(I) = \sum_{t=1}^{T} \mathbb{E}[\tilde{r}_t]$. Thus, the key point is to find the largest possible γ such that the policy π_γ is feasible for all problem setups I, i.e., the random variables $\tilde{X}_{t,\gamma}$ are well defined for each t.

We now find such a γ. For any a and b, denote $\mu_{t,\gamma}(a, b] = P(a < \tilde{X}_{t,\gamma} \leq b)$ assuming $\tilde{X}_{t,\gamma}$ is well defined. A key observation of the Best-fit Magician is that an arriving query with a realized size d_t gets accepted in the previously empty sample path of $\tilde{X}_{t-1,\gamma}$ only if there is less than a γ-measure of sample paths with utilization in $(0, 1 - d_t]$. Then, we can establish an *invariant* that upper-bounds the measure of sample paths with utilization in $(0, b]$ by a decreasing exponential function of the measure with utilization in $(b, 1 - b]$. Our invariant holds for all $b \in (0, 1/2]$, at all times t.

Lemma 2.8.1. *For any $0 < b \leq \frac{1}{2}$ and any $0 < \gamma < 1$ such that $\tilde{X}_{t,\gamma}$ is well-defined, the inequality*

$$\frac{1}{\gamma} \cdot \mu_{t,\gamma}(0, b] \leq \exp(-\frac{1}{\gamma} \cdot \mu_{t,\gamma}(b, 1 - b])$$

holds for all $t = 0, 1, \ldots, T$.

For a fixed t, assume that the random variable $\tilde{X}_{t,\gamma}$ is well defined. Then, given the invariant established in Lemma 2.8.1, we can lower-bound the measure of zero-utilization sample paths,

i.e., $P(\tilde{X}_{t,\gamma} = 0)$, by the constraint

$$\sum_{\tau=1}^{t} \sum_{(r_\tau, d_\tau)} d_\tau \cdot p(r_\tau, d_\tau) = \sum_{\tau=1}^{t} \sum_{(r_\tau, d_\tau)} d_\tau \cdot P((\tilde{r}_\tau, \tilde{d}_\tau) = (r_\tau, d_\tau))$$
$$= \sum_{\tau=1}^{t} \mathbb{E}_{(\tilde{r}_\tau, \tilde{d}_\tau) \sim F_\tau}[\tilde{d}_\tau] \leq 1,$$

where the last inequality holds from our restriction stated at the beginning that $\sum_{t=1}^{T} \mathbb{E}[\tilde{d}_t] \leq 1$. Note that invariant in Lemma 2.8.1 enables us a way to upper-bound the measure of "bad" sample paths with utilization within $(0, b]$ by the measure of sample paths with utilization within $(b, 1-b]$ for some $0 < b < \frac{1}{2}$. Therefore, on the remaining sample paths the utilization must be either 0 or greater than $1-b$, which are good cases for us. From this, we show that a γ as large as $\frac{1}{3+e^{-2}} \approx 0.319$ allows for a γ-measure of sample paths to have zero utilization at time t, which implies that the random variable $\tilde{X}_{t+1,\gamma}$ is well defined. We iteratively apply the above arguments for each $t = 1$ up to $t = T$, and hence we prove the feasibility of our Best-fit Magician policy.

Theorem 2.9. *When $\gamma = \frac{1}{3+e^{-2}}$, the threshold policy π_γ is feasible and has a competitive ratio at least $\frac{1}{3+e^{-2}}$.*

Finally, we show that the guarantee $\gamma = \frac{1}{3+e^{-2}}$ is tight. From Lemma 2.0.2, by bounding the optimal value of Dual$(\boldsymbol{p}, \boldsymbol{D})$ when $\boldsymbol{p}$ and $\boldsymbol{D}$ (deterministic size) are specified as follows,

$$(p_1, d_1) = (1, \epsilon), \quad (p_t, d_t) = \left(\frac{1-2\epsilon}{(T-2)(\frac{1}{2}+\epsilon)}, \frac{1}{2}+\epsilon\right) \text{ for all } 2 \leq t \leq T-1 \text{ and } (p_T, d_T) = (\epsilon, 1) \tag{2.7}$$

for some $\epsilon > 0$, we obtain the following upper bound of the guarantee of any online algorithm.

Theorem 2.10. *For any feasible online policy π, it holds that* $\inf_I \frac{\mathbb{E}_{H\sim I}[V^\pi(H)]}{\mathsf{ExAnte}(I)} \leq \frac{1}{3+e^{-2}}$.

2.5 Improvement in the Unit-density Special Case

In this section, we consider the unit-density special case of our online knapsack problem where $\tilde{r}_t = \tilde{d}_t$ for each t. Then we can suppress the notation $(\tilde{r}_t, \tilde{d}_t)$ and simply use $\tilde{d}_t$ to denote the size and the reward of query t. We modify our previous Best-fit Magician policy to obtain an improved guarantee. Note that now we have the LP upper bound $\mathsf{ExAnte}(I) = \sum_{t=1}^{T} \mathbb{E}[\tilde{r}_t] = \sum_{t=1}^{T} \mathbb{E}[\tilde{d}_t]$. To maximize the total collected reward, it is enough for us to maximize the expected capacity utilization.

We now present our policy in Algorithm 2, which is based on a sequence of probabilities, denoted by $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)$ and satisfying $1 \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_T \geq 0$. Note that if the sequence $\boldsymbol{\gamma}$ is uniform, i.e., $\gamma_1 = \cdots = \gamma_T$, then the modified Best-fit Magician policy in Algorithm 2 is identical to Algorithm 1. We now discuss how to determine the sequence $\boldsymbol{\gamma}$ such that Algorithm 2 is feasible. For any a and b, we denote $\mu_{t,\boldsymbol{\gamma}}(a, b] = P(a < \tilde{X}_{t,\boldsymbol{\gamma}} \leq b)$, where $\tilde{X}_{t,\boldsymbol{\gamma}}$ denotes the

Algorithm 2 Modified Best-fit Magician policy for unit-density special case

1: For a fixed sequence $\boldsymbol{\gamma}$ satisfying $1 \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_T \geq 0$, we initialize $\tilde{X}_{0,\boldsymbol{\gamma}}$ as a random variable taking value 0 with probability 1.
2: **For** $t = 1, 2, \ldots, T$, we do the following:
3: For each realization d_t of $\tilde{d}_t$, we denote $p(d_t) = P(\tilde{d}_t = d_t)$. Then, we denote a threshold $\eta_{t,\boldsymbol{\gamma}}(d_t)$ satisfying

$$P(\eta_{t,\boldsymbol{\gamma}}(d_t) < \tilde{X}_{t-1,\boldsymbol{\gamma}} \leq 1 - d_t) \leq \gamma_t \leq P(\eta_{t,\boldsymbol{\gamma}}(d_t) \leq \tilde{X}_{t-1,\boldsymbol{\gamma}} \leq 1 - d_t). \tag{2.8}$$

4: Denote by $\hat{d}_t$ the realization of query t and by X_{t-1} the consumed capacity at the end of period $t-1$. Then we serve query t if $\eta_{t,\boldsymbol{\gamma}}(\hat{d}_t) < X_{t-1} \leq 1 - \hat{d}_t$; we serve query t with probability $\frac{\gamma_t - P(\eta_{t,\boldsymbol{\gamma}}(\hat{d}_t) < \tilde{X}_{t-1,\boldsymbol{\gamma}} \leq 1-\hat{d}_t)}{P(\eta_{t,\boldsymbol{\gamma}}(\hat{d}_t) = \tilde{X}_{t-1,\boldsymbol{\gamma}})}$ if $X_{t-1} = \eta_{t,\boldsymbol{\gamma}}(\hat{d}_t)$.
5: Based on $\tilde{X}_{t-1,\boldsymbol{\gamma}}$, for each realization d_t of $\tilde{d}_t$ and each point $x \in (\eta_{t,\boldsymbol{\gamma}}(d_t), 1 - d_t]$, we move the $p(d_t) \cdot P(\tilde{X}_{t-1,\boldsymbol{\gamma}} = x)$ measure of probability mass from point x to the new point $x + d_t$ and when $x = \eta_{t,\boldsymbol{\gamma}}(d_t)$, we move the $p(d_t) \cdot (\gamma_t - P(\eta_{t,\boldsymbol{\gamma}}(d_t) < \tilde{X}_{t-1,\boldsymbol{\gamma}} \leq 1 - d_t))$ measure of probability mass from point x to the new point $x + d_t$. We obtain a new distribution and we denote by $\tilde{X}_{t,\boldsymbol{\gamma}}$ a random variable with this distribution.

distribution of capacity utilization at the end of period t in Algorithm 2. Then, we generalize

Lemma 2.8.1 from the uniform sequence to any sequence $\boldsymbol{\gamma}$ satisfying $1 \geq \gamma_1 \geq \cdots \geq \gamma_T \geq 0$.

Lemma 2.10.1. *For any $0 < b \leq \frac{1}{2}$ and any sequence $\boldsymbol{\gamma}$ satisfying $1 \geq \gamma_1 \geq \cdots \geq \gamma_T \geq 0$ such that $\tilde{X}_{t,\boldsymbol{\gamma}}$ is well defined, the inequality*

$$\frac{1}{\gamma_1} \cdot \mu_{t,\boldsymbol{\gamma}}(0, b] \leq \exp(-\frac{1}{\gamma_1} \cdot \mu_{t,\boldsymbol{\gamma}}(b, 1-b])$$

holds for any $t = 1, \ldots, T$.

Using Lemma 2.10.1, we can obtain the following result, which will finally lead to our choice of the feasible sequence $\boldsymbol{\gamma}$ and the guarantee of our algorithm.

Theorem 2.11. *For any t, denote $\psi_t = \mathbb{E}_{\tilde{d}_t \sim F_t}[\tilde{d}_t]$. Then for any sequence $\boldsymbol{\gamma}$ satisfying $1 \geq \gamma_1 \geq \cdots \geq \gamma_T \geq 0$ such that $\tilde{X}_{t,\boldsymbol{\gamma}}$ is well defined, the following inequality holds for each $t = 1, \ldots, T$,*

$$P(\tilde{X}_{t,\boldsymbol{\gamma}} = 0) \geq \min\{1 - \gamma_1 - \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau,\ \ 1 - 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau - \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)\}. \tag{2.9}$$

Note that for each t, if the random variables $\tilde{X}_{\tau,\boldsymbol{\gamma}}$ are well defined for each $\tau \leq t$, and γ_{t+1} satisfies

$$0 \leq \gamma_{t+1} \leq \min\{1 - \gamma_1 - \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau,\ \ 1 - 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau - \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)\}, \tag{2.10}$$

then (2.9) implies that $P(\tilde{X}_{t,\boldsymbol{\gamma}} = 0) \geq \gamma_{t+1}$. Thus, we know that there always exists a threshold $\eta_{t+1,\boldsymbol{\gamma}}(d_t)$ such that (2.8) holds (since it can be set to 0), and the random variable $\tilde{X}_{t+1,\boldsymbol{\gamma}}$ is well defined. We apply the above argument iteratively for each $t = 1$ up to $t = T$. In this way, we conclude that a sufficient condition for the sequence $\boldsymbol{\gamma}$ being feasible is that (2.10) holds for each t.

Note that the expected utilization of Algorithm 2 is $\sum_{t=1}^{T} \gamma_t \cdot \psi_t$. The above analysis implies

that we can focus on solving the following optimization problem to determine the sequence $\boldsymbol{\gamma}$:

$$\begin{aligned}
\mathrm{OP}(\boldsymbol{\psi}) := \max \quad & \sum_{t=1}^{T} \gamma_t \cdot \psi_t \\
\text{s.t.} \quad & \gamma_{t+1} \le 1 - \gamma_1 - \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau, \quad \forall t = 1, \dots, T-1 \\
& \gamma_{t+1} \le 1 - 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau - \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau), \quad \forall t = 1, \dots, T-1 \\
& 1 \ge \gamma_1 \ge \cdots \ge \gamma_T \ge 0,
\end{aligned} \tag{2.11}$$

where $\boldsymbol{\psi} = (\psi_1, \dots, \psi_T)$. It holds that $\sum_{t=1}^{T} \psi_t \le 1$. Our solution to $\mathrm{OP}(\boldsymbol{\psi})$ can be obtained from the following function over the interval $[0, 1]$, the value of which is iteratively computed for a $\gamma_0 \in (0, 1)$:

$$\begin{aligned}
h_{\gamma_0}(0) =& \gamma_0 \\
h_{\gamma_0}(t) =& \min \left\{ \lim_{\tau \to t-} h_{\gamma_0}(\tau),\ 1 - \gamma_0 - \int_{\tau=0}^{t} h_{\gamma_0}(\tau) d\tau, 1 - 2 \cdot \int_{\tau=0}^{t} h_{\gamma_0}(\tau) d\tau - \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot \int_{\tau=0}^{t} h_{\gamma_0}(\tau) d\tau) \right\}.
\end{aligned} \tag{2.12}$$

It is easy to see that the function $h_{\gamma_0}(\cdot)$ is non-increasing and non-negative over $[0, 1]$ as long as $0 < \gamma_0 < 1$. Thus, the function $h_{\gamma_0}(\cdot)$ specifies a feasible solution to $\mathrm{OP}(\boldsymbol{\psi})$ when each component of $\boldsymbol{\psi}$ is infinitesimally small and $T \to \infty$, where $h_{\gamma_0}(t)$ corresponds to γ_t for each $t \in [0, 1]$.

We now show that for arbitrary $\boldsymbol{\psi}$ satisfying $\sum_{t=1}^{T} \psi_t \le 1$, we can still construct a feasible solution to $\mathrm{OP}(\boldsymbol{\psi})$ based on the function $h_{\gamma_0}(\cdot)$ for each fixed $0 < \gamma_0 < 1$. Specifically, we define a set of indices $0 = k_0 \le k_1 \le \cdots \le k_T \le 1$ such that $\psi_t = k_t - k_{t-1}$ for each $t = 1, \dots, T$. Then, we define

$$\hat{\gamma}_t = \frac{\int_{\tau=k_{t-1}}^{k_t} h_{\gamma_0}(\tau) d\tau}{k_t - k_{t-1}} = \frac{\int_{\tau=k_{t-1}}^{k_t} h_{\gamma_0}(\tau) d\tau}{\psi_t} \tag{2.13}$$

for each $t = 1, \dots, T$. We show in the following lemma that $\{\hat{\gamma}_t\}_{t=1}^{T}$ is a feasible solution to $\mathrm{OP}(\boldsymbol{\psi})$.

Lemma 2.11.1. *For each fixed $0 < \gamma_0 < 1$, let $h_{\gamma_0}(\cdot)$ be the function defined in (2.12). Then, for*

any $\boldsymbol{\psi}$*, the solution* $\{\hat{\gamma}_t\}_{t=1}^T$ *is a feasible solution to OP*($\boldsymbol{\psi}$)*, where* $\hat{\gamma}_t$ *is as defined in* (2.13) *for each* $t = 1, \ldots, T$.

Note that for each $\boldsymbol{\psi}$ satisfying $\sum_{t=1}^T \psi_t \leq 1$, if $\{\hat{\gamma}_t\}_{t=1}^T$ is constructed according to (2.13), then it is easy to see that $\sum_{t=1}^T \hat{\gamma}_t \cdot \psi_t = \int_{t=0}^{k_T} h_{\gamma_0}(t)dt$ and thus the guarantee of our policy in Algorithm 2 based on the sequence $\{\hat{\gamma}_t\}_{t=1}^T$ is $\frac{\int_{t=0}^{k_T} h_{\gamma_0}(t)dt}{k_T}$ for some $k_T \in (0, 1]$, where k_T depends on the setup $\boldsymbol{\psi}$. Since the function $h_{\gamma_0}(\cdot)$ is non-increasing and non-negative over $[0, 1]$ as long as $0 < \gamma_0 < 1$, we know that the worst-case setup corresponds to $k_T = 1$, i.e., $\sum_{t=1}^T \psi_t = 1$. Therefore, it is enough to focus on solving the following problem: $\max_{0<\gamma_0<1} \int_{t=0}^1 h_{\gamma_0}(t)dt$ to obtain the guarantee of our policy. Numerically, we can show that when $\gamma_0 \approx 0.3977$, the above optimization problem reaches its maximum, which is 0.3557. We conclude that the guarantee of our policy is 0.3557.

3 A New Framework Beyond LP-based Guarantees

In this chapter, we present details for our new framework to obtain guarantees beyond the LP-based guarantees. In Section 3.1, we present the notations and the problem classes that we will consider. In Section 3.2, we present the formulation of our new framework and interpret it as a new "Type Coverage" problem. In Section 3.3, we apply our framework to the general non IID setting where the reward distribution of each query can be different and obtain new results. The more restrictive IID setting is considered in Section 3.4 and stronger results can be obtained. All the formal proofs are deferred to the appendix, however, the proof techniques are always sketched.

3.1 Problem Formulation

We consider the prophet inequality problem where k out of $T > k$ queries can be accepted. That is, initially there are k *slots*, and T queries arrive in order $t = 1, \ldots, T$, each with a *valuation* $R_t \geqslant 0$ that is revealed only when query t arrives. One must then immediately decide whether to use a slot to accept query t, or to reject query t forever, with this decision being irrevocable. Once all k slots have been used, no further queries can be accepted. The valuations R_t are drawn independently at random from known distributions. The objective is to make accept/reject decisions

on-the-fly, in a way that maximizes the expected sum of valuations of accepted queries.

We assume that each R_t is drawn from a discrete distribution, input as follows. There is a universe of m possible valuations, sorted in the order $r_1 \geqslant \cdots \geqslant r_m \geqslant 0$. For each query t, we let $p_{tj} \geqslant 0$ denote the probability that their valuation R_t realizes to r_j, for all $j = 1, \ldots, m$, with $\sum_{j=1}^{m} p_{tj} = 1$. We refer to index j as the *type* of a query, with smaller indices j said to be *better*. For simplicity, we also assume that queries arrive in exactly the order $t = 1, \ldots, T$, which is known in advance. Although many of the algorithms we discuss hold under certain adversarial manipulations of the arrival order, we do not attempt to make such distinctions comprehensively.

Definition 3.1 (Instance, IID vs. Non-IID)**.** An *instance* I of the prophet inequality problem is defined by the number of slots k, queries T, types m, the valuations $r_1, \ldots, r_m$, and the probability vectors $(p_{1j})_{j=1}^{m}, \ldots, (p_{nj})_{j=1}^{m}$ under the fixed arrival order $t = 1, \ldots, T$. We let $\mathcal{I}_{k,T}$ denote the class of all instances with k slots and T agents, with $\mathcal{I}_k := \bigcup_{T=1}^{\infty} \mathcal{I}_{k,T}$. If p_{tj} is identical to some p_j across all queries t, for each type j, then we say that the instance is *IID* (independent and identically distributed), and we let $\mathcal{I}_{k,T}^{\text{IID}}$ denote the class of all IID instances with k slots and T queries, with $\mathcal{I}_k^{\text{IID}} := \bigcup_{T=1}^{\infty} \mathcal{I}_{k,T}^{\text{IID}}$. We sometimes refer to general instances as *non-IID*.

We are interested in tight guarantees relative to both the prophet and ex-ante relaxation, for both the classes of IID and non-IID instances with a particular number of slots k. That is, we are interested in the values of $\alpha = \inf_{I \in \mathcal{I}} \frac{\mathsf{DP}(I)}{\mathsf{Proph}(I)}$ and $\alpha = \inf_{I \in \mathcal{I}} \frac{\mathsf{DP}(I)}{\mathsf{ExAnte}(I)}$ when $\mathcal{I}$ can be $\mathcal{I}_k$ or $\mathcal{I}_k^{\text{IID}}$ for some k, which will affect the value of α. Many of our results also imply tight guarantees for more granular classes of instances, e.g. $\mathcal{I} = \mathcal{I}_{k,T}$ or $\mathcal{I}^{\text{IID}} = \mathcal{I}_{k,T}$, which restrict to having exactly T agents. Guarantees relative to the ex-ante relaxation are worse than those relative to the prophet, and guarantees are also be worse for larger classes of instances (e.g. non-IID instead of IID). Finally, we sometimes consider the following subclass of policies that are not as powerful as the optimal DP, which would also make guarantees worse.

Definition 3.2 (Static Threshold Policies)**.** A *static threshold* policy accepts the first k agents to

arrive who have valuation at least r_J, or equivalently have a type j with $j \leqslant J$, for some fixed index $J \in [m]$. A static threshold policy is also allowed to set a tie-break probability $\rho \in (0, 1]$, where each agent of type exactly J is accepted (while slots remain) according to an independent coin flip of probability ρ.

We emphasize that a static threshold policy is not allowed to change J on-the-fly based on the remaining number of slots or agents, making them less powerful than the optimal DP. Static policies of this type have been previously studied in Ehsani et al. (2018); Chawla et al. (2020); Arnosti and Ma (2021), who note that the tie-break probability is unnecessary (i.e. one can always set $\rho = 1$) under alternative models where the valuation distributions are continuous.

We further distinguish between *oblivious* static threshold algorithms that must set J, ρ without knowing the cardinal values of $r_1, \ldots, r_m$ (but knowing m and $\{p_{tj} : t \in [T], j \in [m]\}$), vs. an algorithm that can set J, ρ optimally with full knowledge of the instance I. This distinction was introduced by Chawla et al. (2020) and the benefits of being oblivious are elaborated on in Arnosti and Ma (2021). As an example of this distinction, a thresholding rule based on the *median* value of $\mathsf{Proph}(I)$ (as proposed in Samuel-Cahn (1984)) is oblivious, but a thresholding rule based on the *mean* value (as proposed in Kleinberg and Weinberg (2012)) is not.

3.2 General Framework and the Type Coverage Problem

The idea of our general framework is to *explicitly formulate the adversary's optimization problem* of minimizing some prophet inequality ratio, e.g. $\mathsf{DP}(I)/\mathsf{Proph}(I)$, over all instances I belonging to some class. This allows us to *directly compute the tight guarantee* α in different settings, without having to separately construct a lower bound (usually based on analyzing a simple algorithm) and hoping that it is possible to construct a matching upper bound. Of course, the adversary's problem can be highly intractable, because it needs to optimize over the space of distributions, and encapsulate the best response from the (possibly restricted) algorithm on each instance I.

Our general framework overcomes this intractability by breaking down the adversary's problem into two stages. Treating the number of slots k and queries T as fixed, we assume that the adversary first optimizes over m and the type distributions $\{p_{tj} : t \in [T], j \in [m]\}$, and then optimizes over the specific valuations $\{r_j : j \in [m]\}$. The inner optimization problem over r_j's can then be formulated as an LP, which encapsulates the algorithm's best response using constraints that are linear when the p_{tj}'s are fixed. The dual of this LP gives rise to a new problem that is expressed solely in terms of the type distributions, whose optimal solution has a nice structure. This allows us to solve the outer optimization problem over type distributions in many settings.

Before proceeding we define some notation that will simplify the formulation of these problems.

Definition 3.3 (Δ_j, G_{tj})**.** Let $\Delta_j := r_j - r_{j+1}$ for all $j \in [m]$, with r_{m+1} understood to be 0. We will equivalently write the adversary's inner optimization problem using decision variables Δ_j. Recalling that $r_1 \geqslant \cdots \geqslant r_m \geqslant 0$, we have $\Delta_j \geqslant 0$ for all j, where Δ_j can be interpreted as the "valuation gain" when going from type $j+1$ to type j. Also, let $G_{tj} := \sum_{j'=1}^{j} p_{tj'}$, the probability that query t's valuation is at least r_j, for all $t \in [T]$ and $j = 0, \ldots, m$. Note that $0 = G_{t0} \leqslant \cdots \leqslant G_{im} = 1$ for all t.

Proposition 3.4. *For any instance I,*

$$\mathsf{Proph}(I) = \sum_{j=1}^{m} \Delta_j \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}] \text{ and } \mathsf{ExAnte}(I) = \sum_{j=1}^{m} \Delta_j \cdot \min\{\sum_{t=1}^{n} G_{tj}, k\}, \tag{3.1}$$

where $\mathrm{Ber}(G_{tj})$ *denotes an independent Bernoulli random variable with success probability* G_{tj}*.*

Letting Q_j denote $\mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$ (resp. $\min\{\sum_{t=1}^{n} G_{tj}, k\}$), Q_j can be interpreted as the expected number of agents of type j or better accepted by the prophet (resp. ex-ante relaxation), which explains the formulas in (3.1).

We are now ready to formulate the adversary's inner problem of minimizing $\mathsf{DP}(I)/\mathsf{Proph}(I)$ or $\mathsf{DP}(I)/\mathsf{ExAnte}(I)$ over decision variables Δ_j. By Theorem 3.4, as long as the type distributions

given by G_{tj} are fixed, both $\mathsf{Proph}(I)$ and $\mathsf{ExAnte}(I)$ can be expressed as linear combinations of Δ_j, which the adversary normalizes to 1. Subject to this, the adversary then tries to minimize $\mathsf{DP}(I)$.

Definition 3.5 (Inner Problem for Minimizing DP)**.** Consider the following linear program, with Q_j set to $\mathbb{E}[\min\{\sum_{t=1}^{T}\mathrm{Ber}(G_{tj}), k\}]$ (resp. $\min\{\sum_{t=1}^{n} G_{tj}, k\}$) for all $j \in [m]$.

$$\min V_1^k \tag{3.2a}$$

$$\text{s.t. } V_t^l = \sum_{j=1}^{m} p_{tj} U_{tj}^l + V_{t+1}^l \qquad \forall t \in [T], l \in [k] \tag{3.2b}$$

$$U_{tj}^l \geqslant \sum_{j'=j}^{m} \Delta_{j'} - (V_{t+1}^l - V_{t+1}^{l-1}) \qquad \forall t \in [T], j \in [m], l \in [k] \tag{3.2c}$$

$$\sum_{j=1}^{m} Q_j \Delta_j = 1 \tag{3.2d}$$

$$\Delta_j, U_{tj}^l \geqslant 0 \qquad \forall t \in [T], j \in [m], l \in [k] \tag{3.2e}$$

In constraint (3.2b), free variable V_t^l denotes the value-to-go of the DP when query t arrives with exactly l slots remaining, with V_t^l understood to be 0 if $t = T+1$ or $l = 0$. Meanwhile, auxiliary variable U_{tj}^l denotes the utility gain when query t realizes to type j with l slots remaining. The utility gain U_{tj}^l is lower-bounded by both 0 and the expression $\sum_{j'=j}^{m} \Delta_{j'} - V_{t+1}^l + V_{t+1}^{l-1}$ in (3.2c), which denotes the immediate gain from accepting query t (who has valuation $r_j = \sum_{j'=j}^{m} \Delta_{j'}$) minus the loss $(V_{t+1}^l - V_{t+1}^{l-1})$ from proceeding to query $t+1$ with $l-1$ instead of l slots remaining. Finally, constraint (3.2d) normalizes the value of $\mathsf{Proph}(I)$ (resp. $\mathsf{ExAnte}(I)$) to 1.

Therefore, in the linear program V_1^k will equal precisely the optimal performance $\mathsf{DP}(I)$ of dynamic programming (see e.g. Puterman, 2014), and hence LP (3.2) correctly describes the adversary's inner problem of minimizing $\mathsf{DP}(I)/\mathsf{Proph}(I)$ (resp. $\mathsf{DP}(I)/\mathsf{ExAnte}(I)$) over all instances I with some given type distributions. We now take the dual of (3.2) to uncover a new problem.

Definition 3.6 (Dual of Inner Problem for Minimizing DP)**.** Defining dual variables x_t^l, y_{tj}^l, θ for

constraints (3.2b),(3.2c),(3.2d) respectively, the following LP is dual to (3.2).

$$\max \theta \tag{3.3a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \sum_{j'=1}^{j} y_{tj'}^{l} \qquad \forall j \in [m] \tag{3.3b}$$

$$y_{tj}^{l} \leqslant p_{tj} x_t^l \qquad \forall t \in [T], j \in [m], l \in [k] \tag{3.3c}$$

$$x_t^l = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l - \sum_{j=1}^{m} (y_{t-1,j}^l - y_{t-1,j}^{l+1}), & t > 1 \end{cases} \qquad \forall t \in [T], l \in [k] \tag{3.3d}$$

$$y_{tj}^{l} \geqslant 0 \qquad \forall t \in [T], j \in [m], l \in [k] \tag{3.3e}$$

Before trying to interpret the LP (3.3), we notice the following structure. An optimal solution will always saturate constraints (3.3c) for better types j before setting $y_{tj'}^{l} > 0$ for types $j' > j$. This allows us to reformulate the LP using collapsed variables $y_t^l = \sum_j y_{tj}^l$, as formalized below.

Lemma 3.6.1 (Dual Simplification). *LP* (3.3) *has the same optimal value as the following LP.*

$$\max \theta \tag{3.4a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\} \qquad \forall j \in [m] \tag{3.4b}$$

$$y_t^l \leqslant x_t^l \qquad \forall t \in [T], l \in [k] \tag{3.4c}$$

$$x_t^l = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l - y_{t-1}^l + y_{t-1}^{l+1}, & t > 1 \end{cases} \qquad \forall t \in [T], l \in [k] \tag{3.4d}$$

$$y_t^l \geqslant 0 \qquad \forall t \in [T], l \in [k] \tag{3.4e}$$

We refer to (3.4) as an LP since the non-linear term $\min\{y_t^l, G_{tj}x_t^l\}$ can easily be represented using an auxiliary variable and linear constraints.

We note that LP (3.4) has the following interpretation as a "Type Coverage" problem. Free variable x_t^l denotes the probability of having exactly l slots remaining when query t arrives, and y_t^l denotes the (unconditional) probability of accepting the query in this state, which must lie in $[0, x_t^l]$ as enforced by (3.3c), (3.3e). Meanwhile, (3.3d) correctly updates the state probabilities x_t^l based on the acceptance probabilities y_j^l, which are understood to be 0 if $l = k+1$. Finally, $\min\{y_t^l, G_{tj}x_t^l\}$ represents the probability of accepting query t with type j or better when there are l slots remaining, which is bottlenecked by $G_{tj}x_t^l$ (the probability of query t having type j or better in state l) and y_t^l (the unconditional probability of accepting query t in state l). Therefore, constraint (3.3b) says that the expected number of queries of type j or better accepted must be at least θ in comparison to Q_j, which is the number of queries of type j or better accepted by the prophet or ex-ante relaxation. The algorithm's guarantee is then given by the maximum θ that can be uniformly achieved across all types j.

We now formalize some more notation and summarize the developments of this section.

Definition 3.7 (**G**, Simplified Duals for DP)**.** Let **G** denote the collective information about the type distributions, which includes m and the values of G_{tj} that must satisfy $G_{t1} \leqslant \cdots \leqslant G_{im} = 1$ for all t. For any such valid **G**, let $\text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$ (resp. $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$) denote the LP (3.4) where Q_j is set to $\mathbb{E}[\min\{\sum_{t=1}^{T}\text{Ber}(G_{tj}), k\}]$ (resp. $\min\{\sum_{t=1}^{n} G_{tj}, k\}$) for all $j \in [m]$.

Theorem 3.8 (Reformulation of Tight Guarantees for DP)**.** *For any fixed k and $T > k$,*

$$\inf_{I\in\mathcal{I}_{k,T}} \frac{\text{DP}(I)}{\text{Proph}(I)} = \inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G}) \qquad \textit{and} \qquad \inf_{I\in\mathcal{I}_{k,T}} \frac{\text{DP}(I)}{\text{ExAnte}(I)} = \inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G}).$$

Put in words, Theorem 3.8 says that the tight guarantee for the optimal DP relative to the prophet (resp. ex-ante relaxation) is given by $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$ (resp. $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$), which are based on our simplified dual formulation and have reduced the adversary's problem to

be only over type distributions $\mathbf{G}$. Before delving into how to solve these problems over $\mathbf{G}$, we develop analogues of Theorem 3.8 in the settings where the algorithm in the primal problem is restricted to static threshold policies. We note that in the IID special case, there is a significant further simplification that expresses the adversary's entire problem as a single semi-infinite LP, which we derive in Section 3.4.

3.2.1. General Framework for (Oblivious) Static Threshold Algorithms. When the algorithm is restricted to static threshold policies, we can similarly formulate an inner primal LP and take its dual to uncover a new problem that depends on the type distributions but not on the specific valuations of agents. In fact, a nice *distinction* emerges in the dual depending on whether the algorithm must set the static threshold while oblivious to the specific valuations.

Definition 3.9 (Oblivious vs. Non-oblivious Static Thresholds)**.** For any instance I, let $\mathsf{ST}(I)$ denote the expected performance of the best static threshold policy on I, whose parameters J, ρ can be set knowing instance I. By contrast, an *oblivious* static threshold (OST) algorithm must set the parameters J, ρ without knowing the specific query valuations in the instance (but knowing everything else, including the type distributions).

The tight guarantee for OST algorithms relative to the prophet (resp. ex-ante relaxation) is defined by the following sequence of optimizations. First, for a fixed k and T, the adversary sets $\mathbf{G}$, which we recall is defined by m and $\{G_{tj} : t \in [T], j \in [m]\}$. Based on $\mathbf{G}$, the algorithm fixes the parameters J, ρ of the static threshold policy to be used. Finally, the adversary sets the valuations, defined by Δ_j, to minimize the policy's performance relative to $\mathsf{Proph}(I)$ (resp. $\mathsf{ExAnte}(I)$).

We first focus on tight guarantees for OST algorithms, which are simpler to capture using our framework. For a fixed $\mathbf{G}$ and parameters J, ρ chosen by the OST, we write the adversary's inner optimization problem over Δ_j.

Definition 3.10 (Inner Problem for Minimizing OST)**.** Consider the following LP, where coeffi-

cient Q_j can be set to either $\mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G_{tj}), k\}]$ or $\min\{\sum_{t=1}^{n} G_{tj}, k\}$ like before.

$$\min V_1^k \tag{3.5a}$$

$$\text{s.t. } V_t^l = \sum_{j<J} p_{tj} U_{tj}^l + p_{tJ}\rho U_{tJ}^l + V_{t+1}^l \qquad \forall t \in [T], l \in [k] \tag{3.5b}$$

$$U_{tj}^l = \sum_{j'=j}^{m} \Delta_{j'} - V_{t+1}^l + V_{t+1}^{l-1} \qquad \forall t \in [T], j \in [m], l \in [k] \tag{3.5c}$$

$$\sum_{j=1}^{m} Q_j \Delta_j = 1 \tag{3.5d}$$

$$\Delta_j \geqslant 0 \qquad \forall j \in [m] \tag{3.5e}$$

Compared to the inner LP (3.2) for the optimal DP algorithm, LP (3.5) differs in two ways. First, U_{tj}^l is now a free variable that could be negative, representing the change in utility when query t takes type j and is *accepted* with l slots remaining, as set in (3.5c). Second, the policy is now *forced to accept* a query with type $j < J$ w.p. 1 and a query with type $j = J$ w.p. ρ, regardless of l and t, as reflected in constraints (3.5b). This allows the adversary to create a smaller objective value in LP (3.5), ultimately yielding a maximization problem for the dual in which the collapsed variable y_t^l (from the simplification in Lemma 3.6.1) must satisfy essentially the same static threshold rule as defined by J and ρ. This is formalized in the definition and theorem below.

Definition 3.11 (Simplified Duals for OST)**.** For any type distributions $\mathbf{G}$ and static threshold policy J, ρ, let $\text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G})$ (resp. $\text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G})$) be identical to LP $\text{innerLP}_{k,T}^{\text{DP}/\text{Proph}}(\mathbf{G})$ (resp. $\text{innerLP}_{k,T}^{\text{DP}/\text{ExAnte}}(\mathbf{G})$), except with the additional constraints

$$y_t^l = (\sum_{j<J} p_{tj} + p_{tJ}\rho) x_t^l = ((1-\rho)G_{t,J-1} + \rho G_{tJ}) x_t^l \qquad \forall t \in [T], l \in [k]. \tag{3.6}$$

Theorem 3.12 (Reformulating the Tight Guarantees for OST)**.** *For any fixed k and $T > k$, the tight guarantee for OST algorithms relative to the prophet (resp. ex-ante relaxation) is equal to*

$\inf_{\mathbf{G}} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G})$ *(resp.* $\inf_{\mathbf{G}} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G})$*).*

We proceed to study tight guarantees for non-oblivious static thresholds. Earlier, the way in which the static threshold restriction directly translated into dual constraint (3.6) was crucially dependent on the fact that in the primal LP (3.5), J and ρ were set before the Δ_j's. We now show that if J and ρ are decided after the Δ_j's, then this translates into the dual algorithm being able to employ an *arbitrary convex combination* of static threshold rules, which can change the dual objective. First we formulate the adversary's inner problem for minimizing $\text{ST}(I)$, which must set Δ_j's such that the performance of *any* static threshold policy defined by J, ρ is poor.

Definition 3.13 (Inner Problem for Minimizing ST)**.** Consider the following LP, where coefficient Q_j can be set to either $\mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G_{tj}), k\}]$ or $\min\{\sum_{t=1}^{n} G_{tj}, k\}$ like before.

$$\min \alpha \tag{3.7a}$$

$$\text{s.t. } V_t^l(J,\rho) = \sum_{j<J} p_{tj} U_{tj}^l(J,\rho) + p_{tJ}\rho U_{tJ}^l(J,\rho) + V_{t+1}^l(J,\rho) \qquad \forall t, l, J \in [m], \rho \in (0,1] \tag{3.7b}$$

$$U_{tj}^l(J,\rho) = \sum_{j'=j}^{m} \Delta_{j'} - V_{t+1}^l(J,\rho) + V_{t+1}^{l-1}(J,\rho) \qquad \forall t, j, l, J \in [m], \rho \in (0,1] \tag{3.7c}$$

$$\alpha \geqslant V_1^k(J,\rho) \qquad \forall J \in [m], \rho \in (0,1] \tag{3.7d}$$

$$\sum_{j=1}^{m} Q_j \Delta_j = 1 \tag{3.7e}$$

$$\Delta_j \geqslant 0 \qquad \forall j \in [m] \tag{3.7f}$$

We note that LP (3.7) is similar to LP (3.5), except a copy of the variables has been created for every possible static threshold J, ρ, all of which must perform no better than α. The simplified dual formulation corresponding to LP (3.7) will be easier to write using the following additional notation.

Definition 3.14 $(\mathbf{x}, \mathbf{y}, \mathcal{P}_T^k)$**.** Let $\mathbf{x} := (x_t^l)_{t\in[T],l\in[k]}$, $\mathbf{y} := (y_t^l)_{t\in[T],l\in[k]}$, and let $\mathcal{P}_T^k$ denote the set of

vectors $(\mathbf{x}, \mathbf{y})$ that satisfy the simplified dual problem's state-updating constraints (3.4c)–(3.4e).

Definition 3.15 (Simplified Duals for ST)**.** For any $\mathbf{G}$, let $\text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G})$ (resp. $\text{innerLP}_{k,T}^{\text{ST/ExAnte}}(\mathbf{G})$) denote the following LP, with coefficient Q_j set to $\mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G_{tj}), k\}]$ (resp. $\min\{\sum_{t=1}^{n} G_{tj}, k\}$) for all $j \in [m]$.

$$\max \theta \tag{3.8a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \int_{J,\rho} \mu(J,\rho) \left(\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l(J,\rho), G_{tj} x_t^l(J,\rho)\} \right) \qquad \forall j \in [m] \tag{3.8b}$$

$$y_t^l(J,\rho) = ((1-\rho)G_{t,J-1} + \rho G_{tJ}) x_t^l(J,\rho) \qquad \forall t, l, J \in [m], \rho \in (0,1] \tag{3.8c}$$

$$(\mathbf{x}(J,\rho), \mathbf{y}(J,\rho)) \in \mathcal{P}_T^k \qquad \forall J \in [m], \rho \in (0,1] \tag{3.8d}$$

$$\int_{J,\rho} \mu(J,\rho) = 1 \tag{3.8e}$$

$$\mu(J,\rho) \geqslant 0 \qquad \forall J \in [m], \rho \in (0,1] \tag{3.8f}$$

We note that there is no benefit for the primal algorithm using a convex combination of static thresholds (its expectation is maximized by choosing the best one), but since the dual problem has to uniformly cover each type j, there can be a benefit.

Theorem 3.16 (Reformulating the Tight Guarantees for ST)**.** *For any fixed k and $T > k$,*

$$\inf_{I \in \mathcal{I}_{k,T}} \frac{\text{ST}(I)}{\text{Proph}(I)} = \inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G}) \qquad \textit{and} \qquad \inf_{I \in \mathcal{I}_{k,T}} \frac{\text{ST}(I)}{\text{ExAnte}(I)} = \inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{ST/ExAnte}}(\mathbf{G}).$$

3.3 General Framework applied to the Non-IID Setting

In this section we study tight guarantees over general non-IID instances, starting with those for the optimal DP. Recall that for any number of slots k and queries $T > k$, we have established in Theorem 3.8 that the tight guarantees relative to the prophet and ex-ante relaxation are given

by $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$ and $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$ respectively. As a recap, by using the notation from Theorem 3.14 that treats $\mathbf{x}, \mathbf{y}$ as vectors, $\text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$ can be rewritten as

$$\begin{aligned}
\text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G}) = \quad & \max \quad \theta && \text{(3.9a)}\\
& \text{s.t.} \quad \theta \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G_{tj}), k\}] \leqslant \sum_{t=1}^{T}\sum_{l=1}^{k} \min\{y_t^l, G_{tj}x_t^l\} \forall j && \text{(3.9b)}\\
& \quad (\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k, && \text{(3.9c)}
\end{aligned}$$

and $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$ can be rewritten as

$$\begin{aligned}
\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G}) = \quad & \max \quad \theta && \text{(3.10a)}\\
& \text{s.t.} \quad \theta \cdot \min\{\sum_{t=1}^{T} G_{tj}, k\} \leqslant \sum_{t=1}^{T}\sum_{l=1}^{k} \min\{y_t^l, G_{tj}x_t^l\} \ \ \forall j \in [m] && \text{(3.10b)}\\
& \quad (\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k && \text{(3.10c)}
\end{aligned}$$

3.3.1. DP/ExAnte **in Non-IID Setting.** In general non-IID setting, we obtain the following result regarding DP/ExAnte.

Theorem 3.17. $\inf_{\mathbf{G}} \ \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$ *is equal to the optimal objective value of the following problem:*

$$\begin{aligned}
\inf_{\substack{\sum_{t\in[T]} g_t \leqslant k \\ 1 \geqslant g_t \geqslant 0 \ \forall t}} \quad & \max \ \theta && \text{(3.11a)}\\
& \text{s.t.} \quad \theta \cdot g_t \leqslant \sum_{l=1}^{k} \min\{y_t^l, g_i x_t^l\} \quad \forall t \in [T] && \text{(3.11b)}\\
& \quad (\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k && \text{(3.11c)}
\end{aligned}$$

The result derived in Theorem 3.17 has a nice interpretation. The variable g_t for each $t \in [T]$ can be interpreted as the marginal probability that agent t got accepted in the Ex-Ante relaxation.

Denote by θ^* the optimal value of LP (3.11) after taking minimum over $\mathbf{g} = (g_1, \ldots, g_T)$. Denote by $\{\theta^*, \mathbf{x}_\mathbf{g}, \mathbf{y}_\mathbf{g}\}$ a feasible solution to LP (3.11) for a fixed $\mathbf{g}$, where the value of $(\mathbf{x}_\mathbf{g}, \mathbf{y}_\mathbf{g})$ depends on $\mathbf{g}$. Then constraint (3.11b) implies that each agent t got accepted by the policy specified by $(\mathbf{x}_\mathbf{g}, \mathbf{y}_\mathbf{g})$ with a probability at least θ^* conditional on being accepted in the Ex-Ante relaxation, for any $\mathbf{g}$. Such an implication corresponds to the definition of θ^*-balancedness online contention resolution scheme (OCRS) in Feldman et al. (2021). Thus, Theorem 3.17 implies that an OCRS achieves the tight guarantee of the DP policy, with respect to the Ex-Ante relaxation. Note that this point has been previously proved in Jiang et al. (2022a). Here, we prove the same result in an alternative way by exploiting the structures of our LP framework $\inf_\mathbf{G}$ $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$.

3.3.2. Optimal Oblivious Static Thresholds in Non-IID Setting. Recall that for any fixed k and $T > k$, we have established in Theorem 3.12 that the tight guarantees for OST algorithms relative to the prophet and ex-ante relaxation are given by $\inf_\mathbf{G} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G})$ and $\inf_\mathbf{G} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G})$ respectively, where the inner LP's correspond to (3.9) and (3.10) respectively but both have the added constraints

$$y_t^l = ((1-\rho)G_{t,J-1} + \rho G_{tJ})x_t^l \qquad \forall t \in [T], l \in [k]. \tag{3.12}$$

The inner LP's for OST's are substantially easier to analyze because under constraints (3.12), the RHS that is common to (3.9b) and (3.10b) can be rewritten as

$$\sum_{t=1}^{T}\sum_{l=1}^{k} \min\{y_t^l, G_{tj}x_t^l\} = \sum_{t=1}^{T} \min\{((1-\rho)G_{t,J-1} + \rho G_{tJ}), G_{tj}\} \sum_{l=1}^{k} x_t^l, \tag{3.13}$$

where term $\min\{((1-\rho)G_{t,J-1} + \rho G_{tJ}), G_{tj}\}$ for each agent t depends on the choices of J, ρ but not on the number of remaining slots l. Moreover, we have the following relationships.

Lemma 3.17.1. *Fix an OST J, ρ and define $\tau_t = (1-\rho)G_{t,J-1} + \rho G_{t,J}$ for all $t \in [T]$. Suppose vectors*

$\mathbf{x}, \mathbf{y}$ *satisfy* (3.12)*, t.e.* $y_t^l = \tau_i x_t^l$ *for all* t *and* l*, as well as* $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k$*. Then for all* $t \in [T]$*, we have*

$$\sum_{l=1}^{k} x_t^l = \Pr[\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k] \qquad \textit{and} \qquad \sum_{t'=1}^{t} \tau_{t'} \sum_{l=1}^{k} x_{t'}^l = \mathbb{E}[\min\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}), k\}].$$

Lemma 3.17.1 enjoys a nice interpretation that x_t^l denotes the probability of having exactly l slots remaining when query t arrives. Note that each query t "clears the bar" for acceptance independently with probability τ_t. A query t is accepted if and only if they clear the bar and there is at least 1 slot remaining when they arrive, with the latter probability given by $\sum_{l=1}^{k} x_t^l$. The second part of Lemma 3.17.1 then follows because the number of queries accepted among $t' = 1, \ldots, t$ is equal to the number of them who clear the bar, truncated by k. Meanwhile, the first part of Lemma 3.17.1 follows because there is a slot remaining for query t if and only if the number of previous queries $t' < t$ who cleared the bar is less than k.

Equipped with Lemma 3.17.1, we are now ready to prove our result that the tight guarantees for OST algorithms relative to the stronger ex-ante benchmark are no worse than relative to the prophet. We first show a lower bound of OST/ExAnte in the following lemma.

Lemma 3.17.2. *For any type distributions* $\mathbf{G}$ *and static threshold policy* J, ρ*, we have*

$$\begin{aligned}
&\mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G}) \\
&\geqslant \min\left\{\Pr\left[\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k}\right\}.
\end{aligned}$$

We then show an upper bound of OST/Proph in the following lemma, which matches the lower bound established in Lemma 3.17.2.

Lemma 3.17.3. *For any fixed type distributions* $\mathbf{G}$*, the value of* $\inf_{\mathbf{G}'} \sup_{J,\rho} \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G}')$

can be at most

$$\sup_{J,\rho} \min \left\{ \Pr\left[\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k} \right\}. \tag{3.14}$$

Note that the Ex-Ante benchmark is a stronger benchmark than the prophet. Combining Lemma 3.17.2 and Lemma 3.17.3, we have the following result.

Theorem 3.18. *For any fixed k and T,*

$$\inf_{\mathbf{G}} \sup_{J,\rho} \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{Proph}}(\mathbf{G}) = \inf_{\mathbf{G}} \sup_{J,\rho} \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{ExAnte}}(\mathbf{G}) \tag{3.15}$$

$$= \inf_{\mathbf{G}} \sup_{J,\rho} \min \left\{ \Pr\left[\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k} \right\}.$$

It is direct to see that for the formula in (3.15), the first term in the min operator is decreasing over the threshold J, ρ, while the second term in the min operator is increasing over the threshold J, ρ. Thus, in order to achieve the supremum, the two terms within the min operator must be equivalent, which yields the following result.

Corollary 3.19. *For any number of slots k and queries $T > k$, the best-possible guarantees for OST algorithms relative to the prophet or ex-ante relaxation are identically equal to*

$$\min \ \alpha \tag{3.16a}$$

$$\text{s.t. } \alpha = \Pr\left[\sum_{t=1}^{T-1} \mathrm{Ber}(q_t) < k\right] = \frac{\mathbb{E}[\min\{\sum_{t=1}^{T-1} \mathrm{Ber}(q_t), k\}]}{k} \tag{3.16b}$$

$$q_t \in [0,1] \quad \forall t \in [T-1] \tag{3.16c}$$

Chawla et al. (2020) show that for a fixed k, the infimum value of problem (3.16) over $T > k$ occurs as $T \to \infty$, and is equal to $\Pr[\mathrm{Pois}(\lambda) < k] = \frac{\mathbb{E}[\min\{\mathrm{Pois}(\lambda), k\}]}{k}$, where λ is the unique real

number that makes these quantities identical. They also show how to achieve this guarantee using an oblivious static threshold algorithm. Our framework shows that their guarantees are tight, regardless of whether one is comparing to the prophet or ex-ante relaxation, and moreover never required explicitly computing its value or constructing a family of counterexamples to establish a matching upper bound!

3.3.3. Oblivious vs. Non-oblivious Static Thresholds in Non-IID Setting. We further study the performances of oblivious static threshold policies versus general non-oblivious static threshold policies in the non-IID setting. We first show that there exists an instance $\mathbf{G}$ such that ST performs better than OST, with respect to both the Ex-Ante benchmark and the prophet.

Lemma 3.19.1. *There exists an instance* $\mathbf{G}$ *such that*

$$\text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G}) > \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G})$$

and

$$\text{innerLP}_{k,T}^{\text{ST/ExAnte}}(\mathbf{G}) > \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G})$$

In order to prove Lemma 3.19.1, we can construct an instance where there are two queries, $T = 2$, four types, $m = 4$, and one slot, $k = 1$. The distribution for the first query is given by $\mathbf{G}_1 = (0, \frac{1}{2}, \frac{1}{2}, 1)$, and the distribution for the second query is given by $\mathbf{G}_2 = (\varepsilon, \varepsilon, 1, 1)$, for some $\varepsilon \to 0$. For this instance, we show that ST/Proph and ST/ExAnte achieve the value of $2/3$, while the values for OST/Proph and OST/ExAnte are no greater than $1/2$.

Therefore the method of Chawla et al. (2020) is not instance-optimal. However, we now show that it is optimal in the worst case, hence their bound is tight even for the more powerful class of ST policies. First we need the following analogue of Lemma 3.17.3 for non-oblivious static thresholds.

Lemma 3.19.2. *For any fixed type distributions* $\mathbf{G}$*, the value of* $\inf_{\mathbf{G}'} \text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G}')$ *can be at*

most the supremum of

$$\min\left\{\int_{J,\rho}\Pr\left[\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ})<k\right]\mu(J,\rho),\int_{J,\rho}\frac{\mathbb{E}[\min\{\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ}),k\}]}{k}\mu(J,\rho)\right\}$$

over all measures μ over $J\in[m]$ and $\rho\in(0,1]$.

In Theorem 3.18, we have shown that both the quantities $\inf_{\mathbf{G}}\sup_{J,\rho}\ \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G})$ and $\inf_{\mathbf{G}}\sup_{J,\rho}\ \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G})$ are identical to the quantity

$$\inf_{\mathbf{G}}\sup_{J,\rho}\ \min\left\{\Pr\left[\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ})<k\right],\frac{\mathbb{E}[\min\{\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ}),k\}]}{k}\right\}. \tag{3.17}$$

The results in Chawla et al. (2020) imply that this infimum is in fact achieved in the simple case where there is only $m=1$ type and $G_{t1}=1$ for all t, in which case (3.17) reduces to

$$\sup_{\rho\in(0,1]}\ \min\left\{\Pr\left[\mathrm{Bin}(T-1,\rho)<k\right],\frac{\mathbb{E}[\min\{\mathrm{Bin}(T-1,\rho),k\}]}{k}\right\}. \tag{3.18}$$

We now show that $\inf_{\mathbf{G}}\mathrm{innerLP}_{k,T}^{\mathsf{ST}/\mathsf{ExAnte}}(\mathbf{G})$ cannot do any better, because even the larger quantity from Lemma 3.19.2, which on this one-type instance reduces to

$$\sup_{\mu:(0,1]\to\mathbb{R}_{\geqslant 0},\int_\rho\mu(\rho)=1}\ \min\left\{\int_\rho\Pr\left[\mathrm{Bin}(T-1,\rho)<k\right]\mu(\rho),\int_\rho\frac{\mathbb{E}[\min\{\mathrm{Bin}(T-1,\rho),k\}]}{k}\mu(\rho)\right\}, \tag{3.19}$$

is actually no larger than (3.18).

Theorem 3.20. *For any fixed T and k, it holds that*

$$\inf_{\mathbf{G}} \ \text{innerLP}^{\text{ST/Proph}}_{k,T}(\mathbf{G}) = \inf_{\mathbf{G}} \ \text{innerLP}^{\text{ST/ExAnte}}_{k,T}(\mathbf{G})$$
$$= \inf_{\mathbf{G}} \sup_{J,\rho} \ \text{innerLP}^{\text{OST}(J,\rho)/\text{Proph}}_{k,T}(\mathbf{G}) = \inf_{\mathbf{G}} \sup_{J,\rho} \ \text{innerLP}^{\text{OST}(J,\rho)/\text{ExAnte}}_{k,T}(\mathbf{G}).$$

3.4 Further Simplified General Framework, applied to the IID Setting

In this section we study tight guarantees for DP/OST/ST algorithms relative to the prophet/ex-ante relaxation, under the restriction that the type distributions $\mathbf{G}$ must be IID. That is, for each type $j \in [m]$, it is imposed that G_{tj} is identically equal to some value G_j across all queries $t \in [T]$.

When this is the case, the key dual constraints from Section 3.3 that compare to the prophet, e.g. (3.9b) in $\text{innerLP}^{\text{DP/Proph}}_{k,T}(\mathbf{G})$, can be rewritten as

$$\theta \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G_j), k\}] \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_j x_t^l\} \qquad \forall j \in [m]. \tag{3.20}$$

where the values G_j no longer depend on t. Consequently, these constraints will always be hardest to satisfy when the IID type distribution becomes infinitely granular, t.e. $G_j = j/m$ for all $j \in [m]$ with $m \to \infty$, simply because there are more constraints. The overall problem of interest, $\inf_{\mathbf{G}} \text{innerLP}^{\text{DP/Proph}}_{k,T}(\mathbf{G})$, can then have its outer optimization problem dropped, since the infimum always arises at the infinitely-granular $\mathbf{G}$. This leads to a drastic reduction where the tight guarantees are now described by a single semi-infinite LP, as formalized below. We note that the same reduction can be made for the dual constraints from before that compare to the

ex-ante relaxation:

$$\theta \cdot \min\{\sum_{t=1}^{T} G_j, k\} \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_j x_t^l\} \qquad \forall j \in [m]. \tag{3.21}$$

Definition 3.21 (Semi-infinite LP's for DP in IID Setting)**.** Let $\text{Bin}(T, q)$ denote a Binomial random variable with T independent trials of success probability q. Consider the semi-infinite families of constraints

$$\theta \cdot \mathbb{E}[\min\{\text{Bin}(T, q), k\}] \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, q x_t^l\} \qquad \forall q \in (0, 1] \tag{3.22}$$

$$\theta \cdot \min\{nq, k\} \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, q x_t^l\} \qquad \forall q \in (0, 1] \tag{3.23}$$

which correspond to the limiting cases of (3.20) and (3.21) as $G_j = j/m = q$ and $m \to \infty$. Then for any k and $T > k$, let $\text{iidLP}_{k,T}^{\text{DP/Proph}}$ (resp. $\text{iidLP}_{k,T}^{\text{DP/ExAnte}}$) denote the semi-infinite LP defined by: maximize θ, subject to constraints (3.22) (resp. (3.23)) and $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k$.

Static threshold policies J, ρ also simplify nicely in this semi-infinite dual problem in the IID setting. Namely, the previous dual constraints (3.12) for a static threshold policy, $y_t^l = ((1 - \rho)G_{t,J-1} + \rho G_{tJ})x_t^l$, by setting $\tau = (1 - \rho)G_{t,J-1} + \rho G_{tJ}$ which is identical across t, can be reduced to

$$y_t^l = \tau x_t^l \quad \forall t \in [T], l \in [k] \tag{3.24}$$

Note that $\tau \in (0, 1]$ is a single number in the IID setting; we no longer need an index $J \in [m]$ combined with a tiebreak probability $\rho \in (0, 1]$.

Definition 3.22 (Semi-infinite LP's for OST in IID Setting)**.** For any k, $T > k$, and fixed static threshold policy defined by τ, let $\text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}}$ (resp. $\text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{ExAnte}}$) denote the semi-infinite LP defined by: maximize θ, subject to (3.22) (resp. (3.23)), constraints (3.24) for a static threshold policy in the IID setting, and $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k$.

Finally, for non-oblivious static thresholds, we define the analogue of (3.8) for the IID setting.

Definition 3.23 (ST in IID Setting)**.** For any k and $T > k$, let $\text{iidLP}_{k,T}^{\text{ST/Proph}}$ (resp. $\text{iidLP}_{k,T}^{\text{ST/ExAnte}}$) denote the following optimization problem, when $Q(q) = \mathbb{E}[\min\{\text{Bin}(T, q), k\}]$ (resp. $Q(q) = \min\{nq, k\}$) for all $q \in (0, 1]$.

$$\max \theta \tag{3.25a}$$

$$\text{s.t. } \theta \cdot Q(q) \leqslant \int_{\tau\in(0,1]} \mu(\tau) \left(\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l(\tau), q x_t^l(\tau)\} \right) \qquad \forall q \in (0, 1] \tag{3.25b}$$

$$y_t^l(\tau) = \tau x_t^l(\tau) \qquad \forall t \in [T], l \in [k], \tau \in (0, 1] \tag{3.25c}$$

$$(\mathbf{x}(\tau), \mathbf{y}(\tau)) \in \mathcal{P}_T^k \qquad \forall \tau \in (0, 1] \tag{3.25d}$$

$$\int_{\tau\in(0,1]} \mu(\tau) = 1 \tag{3.25e}$$

$$\mu(\tau) \geqslant 0 \qquad \forall \tau \in (0, 1] \tag{3.25f}$$

We now formalize that all of these are the correct formulations, which compute tight guarantees for DP/OST/ST algorithms relative to the prophet/ex-ante relaxation in the IID special case.

Theorem 3.24 (Reformulations in IID Setting)**.** *For any k and $T > k$, when type distributions* $\mathbf{G}$

are constrained to be IID, the adversary's optimization problems over $\mathbf{G}$ *can be reformulated as*

$$\begin{aligned}
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G}) && = \text{iidLP}_{k,T}^{\text{DP/Proph}} \\
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G}) && = \text{iidLP}_{k,T}^{\text{DP/ExAnte}} \\
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G}) && = \sup_{\tau} \text{innerLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}} \\
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G}) && = \sup_{\tau} \text{innerLP}_{k,T}^{\text{OST}(\tau)/\text{ExAnte}} \\
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G}) && = \text{iidLP}_{k,T}^{\text{ST/Proph}} \\
&\inf_{\mathbf{G}:G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{ST/ExAnte}}(\mathbf{G}) && = \text{iidLP}_{k,T}^{\text{ST/ExAnte}}
\end{aligned}$$

3.4.1. Equivalence of DP/ExAnte**,** OST/Proph**, and** OST/ExAnte **in IID Setting.** Similar to before, OST's are easier to analyze because under constraints (3.24), the RHS that is common to (3.22) and (3.23) can be rewritten as

$$\sum_{t=1}^{T}\sum_{l=1}^{k} \min\{y_t^l, qx_t^l\} = \min\{\tau, q\} \sum_{t=1}^{T}\sum_{l=1}^{k} x_t^l, \tag{3.26}$$

where now term $\min\{\tau, q\}$ depends only on the choice of τ and not on query t nor the number of remaining slots l. Moreover, the analogues of the relationships in Lemma 3.17.1 are that assuming $y_t^l = \tau x_t^l$ for all t and l, for all $t \in [T]$, we have

$$\sum_{l=1}^{k} x_t^l = \Pr[\text{Bin}(t-1,\tau) < k] \qquad \text{and} \qquad \tau \sum_{t'=1}^{t}\sum_{l=1}^{k} x_{t'}^l = \mathbb{E}[\min\{\text{Bin}(t,\tau), k\}]. \tag{3.27}$$

This allows us to prove the following theorem. Although this result is well-known in the literature, we emphasize that our framework simultaneously obtain both the lower bound and the upper bound of the ratios, without explicitly constructing a counterexample to show the upper bound and constructing a policy to show the lower bound.

Theorem 3.25. *For any fixed k and T,*

$$\mathrm{iidLP}_{k,T}^{\mathsf{DP/ExAnte}} = \sup_{\tau} \ \mathrm{iidLP}_{k,T}^{\mathsf{OST}(\tau)/\mathsf{Proph}} = \sup_{\tau} \ \mathrm{iidLP}_{k,T}^{\mathsf{OST}(\tau)/\mathsf{ExAnte}} = \frac{\mathbb{E}[\min\{\mathrm{Bin}(T, k/T), k\}]}{k}.$$

3.4.2. Equivalence of Oblivious and Non-oblivious Static Thresholds in IID Setting. We now show that even the best static threshold performance $\mathsf{ST}(I)$ knowing the full instance I cannot beat the quantities in Theorem 3.25. In fact, we prove a stronger result which was not true in the non-IID setting—$\mathsf{ST}(I)$ is no better than oblivious static thresholds on *any* instance.

To prove this fact, we show that in the ST dual problem (3.25), one never benefits from using a convex combination of thresholds instead of a single threshold. It suffices to show that given any two thresholds $\underline{\tau}, \overline{\tau}$ with $\underline{\tau} < \overline{\tau}$, there exists a τ lying between them which contributes more to the RHS of (3.25b) than the average of $\underline{\tau}$ and $\overline{\tau}$, *simultaneously* for every $q \in (0, 1]$. This is formalized in the theorem below.

Theorem 3.26. *Given any $\underline{\tau}, \overline{\tau}$ with $\underline{\tau} < \overline{\tau}$, there exists a $\tau \in [\underline{\tau}, \overline{\tau}]$ such that*

$$\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l(\tau), q x_t^l(\tau)\} \geqslant \frac{1}{2} \left(\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l(\underline{\tau}), q x_t^l(\underline{\tau})\} + \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l(\overline{\tau}), q x_t^l(\overline{\tau})\} \right) \quad \forall q \in (0, 1]. \tag{3.28}$$

The main idea involves setting τ so that $\mathbb{E}[\min\{\mathrm{Bin}(T, \tau), k\}] = (\mathbb{E}[\min\{\mathrm{Bin}(T, \underline{\tau}), k\}] + \mathbb{E}[\min\{\mathrm{Bin}(T, \overline{\tau}), k\}])/2$, and showing through a careful sequence of stochastic comparisons that this implies

$$\sum_{t=1}^{T} \Pr[\mathrm{Bin}(t-1, \tau) < k] \geqslant \frac{1}{2} \left(\sum_{t=1}^{T} \Pr[\mathrm{Bin}(t-1, \underline{\tau}) < k] + \sum_{t=1}^{T} \Pr[\mathrm{Bin}(t-1, \overline{\tau}) < k] \right). \tag{3.29}$$

That is, if τ is set so that the number of queries accepted is the mean of that for $\underline{\tau}$ and $\overline{\tau}$, then its average probability of having a slot available across $t = 1, \ldots, T$ can only be higher than the

mean of that for $\underline{\tau}$ and $\overline{\tau}$. We note that simply setting $\tau = (\underline{\tau} + \overline{\tau})/2$ does not work, which can be seen through a simple example where $k = 1$, $T = 3$, $\underline{\tau} = 1/3$, $\overline{\tau} = 1$. Then the LHS of (3.29) when $\tau = 2/3$ is $1 + 1/3 + 1/9 = 13/9$. Meanwhile, the RHS of (3.29) is $\frac{(1+2/3+4/9)+1}{2} = 14/9$, and hence setting $\tau = (\underline{\tau} + \overline{\tau})/2$ would not suffice.

Theorem 3.26 shows that when designing the convex combination of thresholds given by $\mu(\tau)$ in problem (3.25), if there is mass on two distinct thresholds $\underline{\tau}, \overline{\tau}$, then it is always better to move them to be a single mass at some intermediate threshold. This means that ultimately it is better to place all the mass at one point, leading to the following theorem.

Corollary 3.27. *It holds that*

$$\text{iidLP}_{k,T}^{\text{ST/Proph}} = \sup_{\tau} \text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}}$$

and

$$\text{iidLP}_{k,T}^{\text{ST/ExAnte}} = \sup_{\tau} \text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{ExAnte}}.$$

We note that the same argument works on a particular instance, defined by the type distribution given by $\{G_j : j \in [m]\}$ that is common across queries (in which case constraints (3.25b) only need to be checked at the points $q = G_j$ for some $j \in [m]$). Therefore, OST's are *instance-optimal* for static threshold policies in the IID setting.

3.4.3. DP/Proph **in IID Setting: A Simplified Derivation of the Tight** $\alpha = 0.745$ **for** $k = 1$**.** We now assume $k = 1$ and show how our framework can be used to recover the tight guarantee 0.745 for the optimal online policy relative to the prophet, by further exploiting the optimality structure of $\text{iidLP}_{1,T}^{\text{DP/Proph}}$. In the following part, we first derive a relaxation of $\text{iidLP}_{1,T}^{\text{DP/Proph}}$, denoted as $\text{LP}_T^{\text{relax}}$, and we obtain an optimal solution of $\text{LP}_T^{\text{relax}}$ with a closed form. We then establish that the relaxation does not improve the objective value, t.e. $\text{iidLP}_{1,T}^{\text{DP/Proph}} = \text{LP}_T^{\text{relax}}$, showing that an optimal solution of $\text{LP}_T^{\text{relax}}$ can be transformed into an optimal solution of $\text{iidLP}_{1,T}^{\text{DP/Proph}}$. Finally, based on the closed form solution of $\text{LP}_T^{\text{relax}}$, we show that the worst case occurs as $T \to \infty$ and

we obtain 0.745 as the tight guarantee. We now elaborate on the key techniques in this three-step derivation.

First step. We rewrite $\text{iidLP}_{1,T}^{\text{DP/Proph}}$ with superscript l omitted:

$$\begin{aligned}
\text{iidLP}_{1,T}^{\text{DP/Proph}} = \quad \max \quad & \theta && \text{(3.30a)}\\
\text{s.t.} \quad & (1-(1-\kappa)^T)\theta \leqslant \sum_{t=1}^{T} \min\{y_t, \kappa(1-\sum_{t'=1}^{t-1} y_{t'})\} \quad \forall \kappa \in (0,1] && \text{(3.30b)}\\
& \sum_{t=1}^{n} y_t \leqslant 1 \\
& y_t \geqslant 0 \quad \forall t \in [T]
\end{aligned}$$

This is a semi-infinite program. The constraint (3.30b) is equivalent to

$$(1-(1-\kappa)^T)\theta \leq \sum_{t\in S} y_t + \kappa \cdot \sum_{t\in[T]\setminus S} (1-\sum_{t'=1}^{t-1} y_{t'}) \qquad \forall \kappa \in [0,1], S \subseteq [T], \tag{3.31}$$

which can be relaxed to

$$(1-(1-\kappa)^T)\theta \leqslant \sum_{t=1}^{I} y_t + \kappa \sum_{t=I+1}^{T} (1-\sum_{t'=1}^{t-1} y_{t'}) \qquad \forall \kappa \in (0,1], I = 0,1\ldots,T. \tag{3.32}$$

We later show that this relaxation does not improve the objective value of $\text{iidLP}_{1,T}^{\text{DP/Proph}}$.

Letting $Y_t = \sum_{t'=1}^{i} y_{t'}$ for all $t \in [T]$, we now consider the optimization problem

$$\begin{aligned}
\max \quad & \theta && \text{(3.33a)}\\
\text{s.t.} \quad & \max_{\kappa\in(0,1]} \left\{ (1-(1-\kappa)^T)\theta - \kappa \sum_{t=I}^{T-1} (1-Y_t) \right\} \leqslant Y_I \quad \forall I = 0,\ldots,T && \text{(3.33b)}\\
& Y_0 \leq 0 \leqslant Y_1 \leqslant \cdots \leqslant Y_T \leqslant 1. && \text{(3.33c)}
\end{aligned}$$

Notice that the LHS of (3.33b) involves solving a concave maximization problem with κ being the decision variable, which has a closed-form solution. Thus the semi-infinite constraint (3.33b)

can be replaced by an equivalent nonlinear but finite-dimensional constraint. Thus, we have obtained a finite-dimensional nonlinear program as a relaxation of $\mathrm{iidLP}_{1,T}^{\mathrm{DP/Proph}}$. This discussion is formalized in Lemma 3.27.1 below. Its proof, which requires a variable substitution of the form $z_I = \frac{1}{T\theta}\sum_{t=I}^{T-1}(1-Y_t)$, is elementary.

Lemma 3.27.1 (Relaxation after Eliminating κ and Substituting Variables)**.** *It holds that* $\mathrm{iidLP}_{1,T}^{\mathrm{DP/Proph}} \leqslant \mathsf{LP}_T^{\mathrm{relax}}$ *where*

$$\begin{aligned}
\mathsf{LP}_T^{\mathrm{relax}} := \quad \max \quad & \theta \\
\text{s.t.} \quad & (T-1)z_I^{T/(T-1)} \leqslant nz_{I+1} + \frac{1}{\theta} - 1 \qquad \forall I = 0,\dots,T-1 \\
& z_0 = 1, z_T = 0 \\
& z_t \in [0,1] \qquad \forall t \in [T-1].
\end{aligned}$$

The optimization problem in Lemma 3.27.1 then has the following structured optimal solution where the z_I's are decreasing from $z_0 = 1$ to $z_T = 0$.

Lemma 3.27.2 (Closed-Form Solution for $\mathsf{LP}_T^{\mathrm{relax}}$)**.** *Denote* $\{\theta, z_I\}_{I=0}^T$ *such that* $z_0 = 1$, $z_T = 0$ *and*

$$z_{I+1} = \frac{T-1}{T} z_I^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T}, \quad \forall I = 0,\dots,T-1$$

Then $\{\theta, z_I\}_{I=0}^T$ *is an optimal solution of* $\mathsf{LP}_T^{\mathrm{relax}}$.

Second step. We must show that the solution constructed in Lemma 3.27.2 can be converted into a feasible solution of $\mathrm{iidLP}_{1,T}^{\mathrm{DP/Proph}}$ with identical objective value. The challenge lies in verifying that the reverse substitution $y_t = T\theta(z_{t+1} - 2z_t + z_{t-1})$ (with the z_t's defined according to Lemma 3.27.2) satisfies all of the constraints in $\mathrm{iidLP}_{1,T}^{\mathrm{DP/Proph}}$, which can be distilled down to showing inequality (3.31). A priori, (3.31) is only satisfied when S takes the form $\{1,\dots,I\}$ (since those are the constraints we kept in the first relaxation (3.32)); in fact (3.31) is even non-obvious

when $\kappa = 0$ and S consists of a singleton t (in which case (3.31) is equivalent to analytically checking that $y_t \geqslant 0$). To streamline the proof of (3.31), we define a set function $f(S)$ that substitutes the pessimal value of κ into (3.31) for each set S, and show this set function to be supermodular. This allows us to ultimately show that it is maximized when S takes the form of an interval $\{1, \ldots, I\}$, for which we already knew by construction that (3.31) is satisfied as equality. This is all formalized in the proof of the lemma below.

Lemma 3.27.3. *It holds that* $\mathrm{iidLP}_{1,T}^{\mathsf{DP/Proph}} = \mathsf{LP}_T^{\mathsf{relax}}$ *for each* $T \geqslant 1$.

Third step. Having established $\mathrm{iidLP}_{1,T}^{\mathsf{DP/Proph}} = \mathsf{LP}_T^{\mathsf{relax}}$, the proof is completed by showing that the objective value of $\mathsf{LP}_T^{\mathsf{relax}}$ is minimized as $T \to \infty$. Although monotonicity in T is difficult to prove in general, we bypass this difficulty by comparing the values of LP_T with LP_{2T}, which our closed-form solution allows us to do.

Lemma 3.27.4. *For any* $T \geqslant 1$, *it holds that* $\mathsf{LP}_T^{\mathsf{relax}} \geq \mathsf{LP}_{2T}^{\mathsf{relax}}$.

Thus, in order to obtain the tight worst case guarantee, it remains to analyze the behavior of the optimal solution $\{\theta, z_I\}_{I=0}^T$ of $\mathsf{LP}_T^{\mathsf{relax}}$ when $T \to \infty$. By Lemma 3.27.2, we have the recursive equation

$$z_{I+1} = \frac{T-1}{T} z_I^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T}, \quad \forall I = 0, \ldots, T-1$$

with $z_0 = 1$ and $z_T = 0$. By treating x as I/T and $H(x)$ as z_I,the recursive equation above motivates the following differential equation,

$$\frac{1}{\theta^*} - 1 = H(x)(\ln H(x) - 1) - H'(x), \qquad \forall x \in [0, 1] \tag{3.34}$$

with the boundary conditions $H(0) = 1$ and $H(1) = 0$, with θ^* being the precise constant that allows these relationships to hold. Then, we have the following result, which is the guarantee for DP/Proph in the IID setting with $k = 1$.

Theorem 3.28. *It holds that* $\inf_T \mathrm{iidLP}_{1,T}^{\mathsf{DP/Proph}} = \lim_{T\to\infty} \mathrm{iidLP}_{1,T}^{\mathsf{DP/Proph}} = \theta^*$*, where* θ^* *is defined in (3.34).*

Note that from the definition of the function $H(x)$ and θ^*, it holds that

$$\int_1^0 \frac{1}{1-\frac{1}{\theta^*}-H(1-\ln H)} dH = 1.$$

This recovers the same integral relationship as in Hill and Kertz (1982); Correa et al. (2017); Liu et al. (2021) which is used to establish the numerical guarantee of $\alpha = \theta^* \approx 0.745$.

Part II

Online Algorithms with Regret Analysis

4 | A Primal-dual Algorithm under Wasserstein-based Non-stationarity

In this chapter, we present our primal-dual algorithm for an online stochastic optimization problem with Wasserstein-based non-stationarity. In Section 4.1, we introduce the notations and present the model formulation. In Section 4.2, we consider a special case where the distributions are given. We present our algorithm and illustrate the main algorithmic idea. In Section 4.3, we present our results for the data-driven setting. We consider the uninformative setting in Section 4.4. All the formal proofs are deferred to the appendix, however, the proof techniques are always sketched.

4.1 Problem Formulation

Consider the following convex optimization problem

$$\begin{aligned} \max \quad & \sum_{t=1}^{T} f_t(\boldsymbol{x}_t) \\ \text{s.t.} \quad & \sum_{t=1}^{T} g_{it}(\boldsymbol{x}_t) \leqslant c_i, \quad i = 1, \ldots, m, \\ & \boldsymbol{x}_t \in \mathcal{X}, \quad t = 1, \ldots, T, \end{aligned} \tag{CP}$$

where the decision variables are $\boldsymbol{x}_t \in \mathcal{X}$ for $t = 1, ..., T$ and $\mathcal{X}$ is a compact convex set in $\mathbb{R}^k$. The function f_t's are functions in the space $\mathcal{F} = \mathcal{F}(\mathcal{X})$ of concave continuous functions and g_{it}'s are functions in the space $\mathcal{G} = \mathcal{G}(\mathcal{X})$ of convex continuous functions, both of which are supported on $\mathcal{X}$. Compactly, we define the vector-value function $\boldsymbol{g}_t(\boldsymbol{x}) = (g_{1t}(\boldsymbol{x}), ..., g_{mt}(\boldsymbol{x}))^\top : \mathbb{R}^k \to \mathbb{R}^m$. Throughout the paper, we use i to index the constraint and t (or sometimes j) to index the decision variables, and we use bold symbols to denote vectors/matrices and normal symbols to denote scalars.

We study the *online stochastic optimization* problem where the functions in the optimization problem (CP) are revealed in an online fashion and one needs to determine the value of decision variables sequentially. Specifically, at each time t, the functions $(f_t, \boldsymbol{g}_t)$ are first revealed, and then the decision maker needs to decide the value of $\boldsymbol{x}_t$. Different from the offline setting, at each time t, we do not have the information of the future part of the optimization problem (from time $t+1$ to T). Given the history $\mathcal{H}_{t-1} = \{f_j, \boldsymbol{g}_j, \boldsymbol{x}_j\}_{j=1}^{t-1}$, the decision of $\boldsymbol{x}_t$ can be expressed as a policy function of the history and the observation at the current time period. That is,

$$\boldsymbol{x}_t = \pi_t(f_t, \boldsymbol{g}_t, \mathcal{H}_{t-1}) \tag{4.1}$$

where the policy function π_t can be time-dependent. We denote the policy $\boldsymbol{\pi} = (\pi_1, ..., \pi_T)$. The decision variables $\boldsymbol{x}_t$'s must conform to the constraints in (CP) throughout the procedure, and the objective is aligned with the maximization objective of the offline problem (CP).

4.1.1. Parameterized Form, Probability Space, and Assumptions. Consider a parametric form of the underlying problem (CP) where the functions $(f_t, \boldsymbol{g}_t)$ are parameterized by a (random) vector $\boldsymbol{\theta}_t \in \Theta \subset \mathbb{R}^l$. Specifically,

$$f_t(\boldsymbol{x}_t) \coloneqq f(\boldsymbol{x}_t; \boldsymbol{\theta}_t), \quad g_{it}(\boldsymbol{x}_t; \boldsymbol{\theta}_t) \coloneqq g_i(\boldsymbol{x}_t; \boldsymbol{\theta}_t)$$

for each $i = 1, ..., m$ and $t = 1, ..., T$. We denote the vector-valued constraint function by $\boldsymbol{g}(\boldsymbol{x}; \boldsymbol{\theta}) =$

$(g_1(\boldsymbol{x};\boldsymbol{\theta}),...,g_m(\boldsymbol{x};\boldsymbol{\theta}))^\top : \mathcal{X} \to \mathbb{R}^m$. Then the problem (CP) can be rewritten as the following parameterized convex program

$$\begin{aligned} \max \quad & \sum_{t=1}^{T} f(\boldsymbol{x}_t;\boldsymbol{\theta}_t) \\ \text{s.t.} \quad & \sum_{t=1}^{T} g_i(\boldsymbol{x}_t;\boldsymbol{\theta}_t) \leqslant c_i, \quad i = 1,...,m, \\ & \boldsymbol{x}_t \in \mathcal{X}, \quad t = 1,...,T, \end{aligned} \tag{PCP}$$

where the decision variables are $(\boldsymbol{x}_1,....,\boldsymbol{x}_T)$. We note that this parametric form (PCP) avoids the complication of dealing with probability measure in function space. It is introduced mainly for presentation purpose, and it will change the nature of the problem. Moreover, we assume the knowledge of f and $\boldsymbol{g}$ a priori. Here and hereafter, we will use (PCP) as the underlying form of the online stochastic optimization problem.

The problem of online stochastic optimization, as its name refers, involves stochasticity on the functions for the underlying optimization problem. The parametric form (PCP) reduces the randomness from the function to the parameters $\boldsymbol{\theta}_t$'s, and therefore the probability measure can be defined in the parameter space of Θ. First, we consider the following distance function between two parameters $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$,

$$\rho(\boldsymbol{\theta},\boldsymbol{\theta}') \coloneqq \sup_{\boldsymbol{x}\in\mathcal{X}} \|(f(\boldsymbol{x};\boldsymbol{\theta}),\boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta})) - (f(\boldsymbol{x};\boldsymbol{\theta}'),\boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta}'))\|_\infty \tag{4.2}$$

where $\|\cdot\|_\infty$ is the supremum norm in $\mathbb{R}^{m+1}$. Without loss of generality, let Θ be a set of class representatives, that is, for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}' \in \Theta$, $\rho(\boldsymbol{\theta},\boldsymbol{\theta}') > 0$. In this way, the parameter space Θ can be viewed as a metric space equipped with metric $\rho(\cdot,\cdot)$. We choose supremum norm because in our model, one *single* action $\boldsymbol{x}_t$ is made at each period t and it can take any arbitrary value in $\mathcal{X}$. As a result, the comparison between $(f(\cdot;\boldsymbol{\theta}),\boldsymbol{g}(\cdot;\boldsymbol{\theta}))$ and $(f(\cdot;\boldsymbol{\theta}'),\boldsymbol{g}(\cdot;\boldsymbol{\theta}'))$ should be made at

each point $\boldsymbol{x}$ and thus the supremum norm is a natural choice for defining $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}')$. In this way, the definition of ρ is based on the vector-valued function $(f, \boldsymbol{g}) : \mathcal{X} \to \mathbb{R}^{m+1}$. Thus it captures the effect of different parameters on the function value rather than the original Euclidean difference in the parameter space. Let $\mathcal{B}_\Theta$ be the smallest σ-algebra in Θ that contains all open subsets (under metric ρ) of Θ. We denote the distribution of $\boldsymbol{\theta}_t$ as $\mathcal{P}_t$ which can be viewed as a probability measure on $(\Theta, \mathcal{B}_\Theta)$.

We now make the following assumptions. Assumption 4.1 (a) and (b) impose boundedness on function f and g_i's. Assumption 4.1 (c) states the ratio between f and g_i is uniformly bounded by q for all $\boldsymbol{x}$ and $\boldsymbol{\theta}$. Intuitively, it tells that for each unit consumption of resource, the maximum amount of revenue earned is upper bounded by q. This condition will mainly be used to give an upper bound on the dual optimal solution. In Assumption 4.1 (d), we assume $\mathcal{P}_t$'s are independent of each other but we do not assume the exact knowledge of them. Also, there can be dependence between components in the vector-value functions $(f, \boldsymbol{g})$. In Assumption 4.1 (e), we require some convexity structure for the underlying functions. In the rest of the paper, this assumption will only be used to ensure that the Lagrangian problem $\max_{\boldsymbol{x}\in\mathcal{X}}\{f(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{p}^\top g(\boldsymbol{x};\boldsymbol{\theta})\}$ can be efficiently solved for any fixed $\boldsymbol{p} \geqslant \boldsymbol{0}$.

Assumption 4.1 (Boundedness and Independence)**.** *We assume*

(a) $|f(\boldsymbol{x};\boldsymbol{\theta})| \leqslant 1$ *for all* $\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta$.

(b) $g_i(\boldsymbol{x};\boldsymbol{\theta}) \in [0,1]$ *for all* $\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta$ *and* $i = 1, ..., m$. *In particular,* $\boldsymbol{0} \in \mathcal{X}$ *and* $g_i(\boldsymbol{0};\boldsymbol{\theta}) = 0$ *for all* $\boldsymbol{\theta} \in \Theta$.

(c) There exists a positive constant q *such that for any* $\boldsymbol{\theta} \in \Theta$ *and each* i, *we have that* $f(\boldsymbol{x};\boldsymbol{\theta}) \leq q \cdot g_i(\boldsymbol{x};\boldsymbol{\theta})$ *holds for any* $\boldsymbol{x} \in \mathcal{X}$ *when* $g_i(\boldsymbol{x};\boldsymbol{\theta}) > 0$.

(d) $\boldsymbol{\theta}_t \sim \mathcal{P}_t$ *and* $\mathcal{P}_t$*'s are independent with each others.*

(e) *The function $f(\boldsymbol{x};\boldsymbol{\theta})$ is concave over $\boldsymbol{x}$ and the function $g_i(\boldsymbol{x};\boldsymbol{\theta})$ is convex over $\boldsymbol{x}$ for any $\boldsymbol{\theta} \in \Theta$ and $i = 1, \ldots, m$.*

4.1.2. Performance Measure. We denote the offline optimal solution of optimization problem (CP) as $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, ..., \boldsymbol{x}_T^*)$, and the offline (online) objective value as R_T^* (R_T). Specifically,

$$R_T^* := \sum_{t=1}^{T} f_t(\boldsymbol{x}_t^*)$$
$$R_T(\boldsymbol{\pi}) := \sum_{t=1}^{T} f_t(\boldsymbol{x}_t).$$

where the online objective value depends on the policy $\boldsymbol{\pi}$. Aligned with general online learning/optimization problems, we focus on minimizing the gap between the online and offline objective values. Specifically, the *optimality gap* is defined as follows:

$$\mathrm{Reg}_T(\mathcal{H}, \boldsymbol{\pi}) := R_T^* - R_T(\boldsymbol{\pi})$$

where the problem profile $\mathcal{H}$ encapsulates the realization of the random parameters, i.e., $\mathcal{H} := (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T)$. Note that R_T^*, $R_T(\pi)$, $\boldsymbol{x}_t^*$ and $\boldsymbol{x}_t$ are all dependent on the problem profile $\mathcal{H}$, but we omit $\mathcal{H}$ in these terms for notation simplicity when there is no ambiguity. We define the performance measure of the online stochastic optimization problem formally as *regret*

$$\mathrm{Reg}_T(\boldsymbol{\pi}) := \max_{\mathcal{P} \in \Xi} \mathbb{E}_{\mathcal{H} \sim \mathcal{P}}[\mathrm{Reg}_T(\mathcal{H}, \boldsymbol{\pi})] \tag{4.3}$$

where $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)$ denotes the probability measure of all time periods and the expectation is taken with respect to the parameter $\boldsymbol{\theta}_t \sim \mathcal{P}_t$; compactly, we write the problem profile $\mathcal{H} \sim \mathcal{P}$. We consider the worst-case regret for all the distribution $\mathcal{P}$ in a certain set Ξ where the set Ξ will be specified in later sections.

The specification of the set Ξ imposes more structure on the distributions of $(\mathcal{P}_1, ..., \mathcal{P}_T)$ and

this is one of the main themes of our paper. In the canonical setting of online stochastic learning problem, all the distributions $\mathcal{P}_t$'s are the same, i.e., $\mathcal{P}_t = \mathcal{P}_0$ for $t = 1, ..., T$. Meanwhile, the adversarial setting of online learning problem refers to the case when $\mathcal{P}_t$'s are adversarially chosen. Our work aims to bridge these two ends of the spectrum with a novel notion of non-stationarity, and to relate the algorithm performance with certain structural property of $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)$. In the same spirit, the work on non-stationary stochastic optimization (Besbes et al., 2015) proposes an elegant notion of non-stationarity called variation budget. Subsequent works consider similar notions in the settings of bandits (Besbes et al., 2014; Russac et al., 2019) and reinforcement learning (Cheung et al., 2019). To the best of our knowledge, all the previous works along this line consider unconstrained setting and thus our work contributes to this line of work in illustrating how the constraints interact with the non-stationarity.

4.2 Known Distribution and Informative Gradient Descent

We begin our discussion with the case when the distributions $\mathcal{P}_t$'s are all known a priori. We use this case to motivate and present our prototypical algorithm – *informative gradient descent* which incorporates the prior information of $\mathcal{P}_t$'s with the online gradient descent algorithm. In the following sections, we will discuss the case when the distributions $\mathcal{P}_t$'s are not known precisely and analyze the algorithm performance accordingly.

4.2.1. Deterministic Upper Bound and Dual Problem. We first introduce the standard deterministic upper bound for the regret benchmark – the "offline" optimum $\mathbb{E}[R_T^*]$. We define the following expectation for a function $u(\boldsymbol{x}; \boldsymbol{\theta}) : \mathcal{X} \to \mathbb{R}$ and a probability measure $\mathcal{P}$ in the parameter space Θ,

$$\mathcal{P}u(\boldsymbol{x}(\boldsymbol{\theta}); \boldsymbol{\theta}) \coloneqq \int_{\boldsymbol{\theta}' \in \Theta} u(\boldsymbol{x}(\boldsymbol{\theta}'); \boldsymbol{\theta}') d\mathcal{P}(\boldsymbol{\theta}')$$

where $\boldsymbol{x}(\boldsymbol{\theta}) : \Theta \to \mathcal{X}$ is a measurable function. Thus $\mathcal{P}u(\cdot)$ can be viewed as a deterministic functional that maps function $\boldsymbol{x}(\boldsymbol{\theta})$ to a real value and it is obtained by taking expectation with respect to the parameter $\boldsymbol{\theta} \sim \mathcal{P}$.

Then, consider the following optimization problem

$$\begin{aligned} R_T^{\mathrm{UB}} = \max \quad & \sum_{t=1}^{T} \mathcal{P}_t f(\boldsymbol{x}_t(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t) \\ \text{s.t.} \quad & \sum_{t=1}^{T} \mathcal{P}_t g_i(\boldsymbol{x}_t(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t) \leqslant c_i, \quad i = 1, ..., m, \\ & \boldsymbol{x}_t(\boldsymbol{\theta}_t) : \Theta \to \mathcal{X} \text{ is a measurable function for } t = 1, ..., T. \end{aligned} \tag{4.4}$$

where $\boldsymbol{\theta}_t$ follows the distribution $\mathcal{P}_t$. The optimization problem (4.4) can be viewed as a convex relaxation of (PCP) where the objective and constraints are all replaced with their expected counterparts. Here $\boldsymbol{x}_{1:T} = (\boldsymbol{x}_1(\boldsymbol{\theta}_1), ..., \boldsymbol{x}_T(\boldsymbol{\theta}_T))$ encapsulates all the primal decision variables. The primal variables are expressed in a function form of $\boldsymbol{\theta}_t$ in that for each different $\boldsymbol{\theta}_t$, we allow a different choice of the primal variables. This is aligned with the "first-observe-then-decide" setting where the decision maker first observes the realization of $\boldsymbol{\theta}_t \sim \mathcal{P}_t$ and then decide the value of $\boldsymbol{x}_t$. As a standard result in literature (Gallego and Van Ryzin, 1994), Lemma 4.1.1 establishes the optimal objective value R_T^{UB} as an upper bound for $\mathbb{E}[R_T^*]$.

Lemma 4.1.1. *It holds that* $R_T^{UB} \geq \mathbb{E}[R_T^*]$.

The deterministic upper bound and the optimization problem (4.4) are commonly used to design algorithms in the literature. To proceed, we introduce the Lagrangian of (4.4),

$$L(\boldsymbol{p}, \boldsymbol{x}_{1:T}) \coloneqq \boldsymbol{c}^\top \boldsymbol{p} + \sum_{t=1}^{T} \mathcal{P}_t \left(f(\boldsymbol{x}_t(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t) - \boldsymbol{p}^\top \boldsymbol{g}(\boldsymbol{x}_t(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t) \right)$$

where $\boldsymbol{\theta}_t$ follows the distribution $\mathcal{P}_t$. The (Lagrange multipliers) vector $\boldsymbol{p} = (p_1, ..., p_m)^\top$ conveys a meaning of shadow price for each budget, and $p_i \geqslant 0$ is the multiplier/dual variable associated

with the i-th constraint. Furthermore, we define the following function based on a point-wise optimization for the primal variables,

$$h(\boldsymbol{p};\boldsymbol{\theta}) \coloneqq \max_{\boldsymbol{x}(\boldsymbol{\theta})\in\mathcal{X}} \left\{ f(\boldsymbol{x}(\boldsymbol{\theta});\boldsymbol{\theta}) - \boldsymbol{p}^\top \boldsymbol{g}(\boldsymbol{x}(\boldsymbol{\theta});\boldsymbol{\theta}) \right\}.$$

Here the point-wise optimization emphasizes that the primal variables can be dependent on (and as a measurable function of) the parameter $\boldsymbol{\theta}_t$. For example, the pricing and assortment decisions can be made upon the observation of the customer type. Then the dual problem of (4.4) becomes

$$\min_{\boldsymbol{p}\geqslant\boldsymbol{0}} L(\boldsymbol{p}) \coloneqq \boldsymbol{c}^\top \boldsymbol{p} + \sum_{t=1}^{T} \mathcal{P}_t h(\boldsymbol{p};\boldsymbol{\theta}_t)$$

where $\boldsymbol{\theta}_t$ follows the distribution $\mathcal{P}_t$ as before.

Let $\boldsymbol{p}^*$ denote an optimal dual solution, i.e.,

$$\boldsymbol{p}^* \in \operatorname{argmin}_{\boldsymbol{p}\geq\boldsymbol{0}} L(\boldsymbol{p}) \tag{4.5}$$

and for each t,

$$\boldsymbol{\gamma}_t \coloneqq \mathcal{P}_t \boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\theta});\boldsymbol{\theta}) \;\text{ where } \boldsymbol{x}^*(\boldsymbol{\theta}) = \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{X}} \{ f(\boldsymbol{x};\boldsymbol{\theta}) - (\boldsymbol{p}^*)^\top \cdot \boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta}) \}. \tag{4.6}$$

Here, $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the associated primal (optimal) solution under the dual optimal solution $\boldsymbol{p}^*$, and $\boldsymbol{\gamma}_t$ can be interpreted as the expected budget consumption in the t-th time period under the optimal primal-dual pair $(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{p}^*)$. Accordingly, for each t, we define

$$L_t(\boldsymbol{p}) \coloneqq \boldsymbol{\gamma}_t^\top \boldsymbol{p} + \mathcal{P}_t h(\boldsymbol{p};\boldsymbol{\theta}).$$

The following proposition states the relation between $L(\cdot)$ and $L_t(\cdot)$ and establishes an upper bound for the benchmark using the dual problem.

Proposition 4.2. *For each $t = 1, ..., T$, it holds that*

$$\boldsymbol{p}^* \in \operatorname{argmin}_{\boldsymbol{p} \geq 0} L_t(\boldsymbol{p})$$

where $\boldsymbol{p}^*$ is defined in (4.5) as one dual optimal solution. Moreover, we have

$$\mathbb{E}[R_T^*] \leq \min_{\boldsymbol{p} \geq 0} \sum_{t=1}^{T} L_t(\boldsymbol{p}) = \sum_{t=1}^{T} \min_{\boldsymbol{p}_t \geq 0} L_t(\boldsymbol{p}_t).$$

4.2.2. Main Algorithm and Regret Analysis. Now we present our main algorithm – *Informative Gradient Descent* – fully described as Algorithm 3. The algorithm is described as a meta algorithm with an input $\boldsymbol{\gamma}$. In the following sections, we will discuss how to apply the algorithm to different settings with different specifications of $\boldsymbol{\gamma}$. When the distributions $\mathcal{P}_t$'s are known, the algorithm is motivated from the dual-based representation in Proposition 4.2 and the input $\boldsymbol{\gamma}$ is accordingly defined by (4.6). Specifically, the algorithm maintains a dual vector/price $\boldsymbol{p}_t$, and at each time t, it performs a stochastic gradient descent update for $\boldsymbol{p}_t$ with respect to the function $L_t(\cdot)$. To see that the expectation of the dual gradient update (A.11) is the gradient with respect to the function $L_t(\cdot)$ evaluated at $\boldsymbol{p}_t$,

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{\gamma}_t\right] &= -\boldsymbol{\gamma}_t + \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}) \\ &= -\frac{\partial}{\partial \boldsymbol{p}} \left(\boldsymbol{\gamma}_t^\top \boldsymbol{p} + \mathcal{P}_t h(\boldsymbol{p}; \boldsymbol{\theta})\right) \Big|_{\boldsymbol{p}=\boldsymbol{p}_t} \end{aligned}$$

where the first line comes from taking expectation with respect to $\boldsymbol{\theta}_t$ and the second line comes from the definition of $\tilde{\boldsymbol{x}}_t$ in the algorithm. At each time t, the primal decision variable $\boldsymbol{x}_t$ is determined based on the dual price $\boldsymbol{p}_t$ and the observation $\boldsymbol{\theta}_t$ jointly, in the same manner as the definition of the function $h(\boldsymbol{p}; \boldsymbol{\theta})$. Assumption 4.1 (e) ensures the optimization problem that defines $h(\boldsymbol{p}; \boldsymbol{\theta})$ can be solved efficiently.

Algorithm 3 Informative Gradient Descent Algorithm (IGD($\boldsymbol{\gamma}$))

Input: parameters $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_T)$.
Initialize the initial dual price $\boldsymbol{p}_1 = \mathbf{0}$ and initial constraint capacity $\boldsymbol{c}_1 = \boldsymbol{c}$
for $t = 1, ..., T$ **do**
    Observe $\boldsymbol{\theta}_t$ and solve

$$\tilde{\boldsymbol{x}}_t = \text{argmax}_{\boldsymbol{x}\in\mathcal{X}}\{f(\boldsymbol{x};\boldsymbol{\theta}_t) - \boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta}_t)\}$$

    where $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{\theta}_t) = (g_1(\boldsymbol{x},\boldsymbol{\theta}_t), ..., g_m(\boldsymbol{x},\boldsymbol{\theta}_t))^\top$
    Set

$$\boldsymbol{x}_t = \begin{cases} \tilde{\boldsymbol{x}}_t, & \text{if } \boldsymbol{c}_t \text{ permits a consumption of } \boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

    Update the dual price

$$\boldsymbol{p}_{t+1} = \left(\boldsymbol{p}_t + \frac{1}{\sqrt{T}}\left(\boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - \boldsymbol{\gamma}_t\right)\right) \vee \mathbf{0} \tag{4.7}$$

    where the element-wise maximum operator $u \vee v = \max\{v, u\}$
    Update the remaining capacity

$$\boldsymbol{c}_{t+1} = \boldsymbol{c}_t - \boldsymbol{g}(\boldsymbol{x}_t;\boldsymbol{\theta}_t)$$

end for
Output: $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$

We now provide an alternative perspective to interpret Algorithm 3 for the case when the distributions are known. Note that by definition (4.6), $\boldsymbol{\gamma}_t$'s represent the "optimal" way to allocate the resource budget over time according to the dual optimal solution $\boldsymbol{p}^*$. Specifically, a larger (resp. smaller) value of $\gamma_{i,t}$, where $\gamma_{i,t}$ denotes the i-th component of $\boldsymbol{\gamma}_t$, indicates that more (resp. less) budget should be allocated to time period t for constraint i. In Algorithm 3, from the update rule (A.11) of the dual variable at time period t, we know that if the budget consumption of constraint i is larger (resp. smaller) than $\gamma_{i,t}$, i.e., $g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) > \gamma_{i,t}$ (resp. $g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) < \gamma_{i,t}$), then we have that $p_{i,t+1} \leqslant p_{i,t}$ (resp. $p_{i,t+1} > p_{i,t}$), where $p_{i,t}$ denotes the i-th component of $\boldsymbol{p}_t$. That is, if more (resp. less) budget is consumed in the earlier periods, then the dual price will be more likely to increase (resp. decrease), and consequently, less (resp. more) budget will be consumed in the future periods. In this way, the dual variable $\boldsymbol{p}_t$ dynamically balances the budget consumption:

it ensures that for each t, the cumulative budget consumption of Algorithm 3 during the first t time periods always stay "close" to the optimal scheme of $\sum_{j=1}^{t} \boldsymbol{\gamma}_j$. Later in Section 4.3, we will show that a dynamic policy that incorporates the resource consumption process into the online decisions is necessary in a non-stationary environment, and any static policy can incur a linear regret even when the underlying non-stationarity intensity is sublinear.

Now we analyze the performance of Algorithm 3 for the known distribution setting. The following lemma says that the dual price vector $\boldsymbol{p}_t$ remains bounded throughout the procedure. Its proof largely relies on Assumption 4.1 (c), and also, the main usage of Assumption 4.1 (c) throughout our analysis is to ensure the boundedness of the dual vector.

Lemma 4.2.1. *Under Assumption 4.1, the dual price vector satisfies* $\|\boldsymbol{p}_t\|_\infty \leq q+1$ *for* $t = 1, 2, \ldots, T$. *Here* $\boldsymbol{p}_t$ *is computed by* (A.11) *of IGD*($\boldsymbol{\gamma}$) *in Algorithm 3 with* $\boldsymbol{\gamma}$ *specified by* (4.6), *and the constant* q *is defined in Assumption 4.1 (c).*

The following theorem builds upon Proposition 4.2 and Lemma 4.2.1 and it states that the regret of Algorithm 3 is upper bounded by $O(\sqrt{T})$.

Theorem 4.3. *Under Assumption 4.1, if we consider the set* $\Xi = \{\mathcal{P} : \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T), \forall \mathcal{P}_1, \ldots, \forall \mathcal{P}_T\}$, *then the regret of IGD*($\boldsymbol{\gamma}$) *in Algorithm 3, with* $\boldsymbol{\gamma}$ *defined by* (4.6), *has the following upper bound*

$$Reg_T(\pi_{IGD}) \leqslant O(\sqrt{T})$$

where π_{IGD} *stands for the policy specified by IGD*($\boldsymbol{\gamma}$) *in Algorithm 3.*

The following proposition 4.4 constructs a problem instance showing that even for a stationary setting where $\mathcal{P}_t$ for each t is identical to each other and known a priori, the lower bound of any online policy is $\tilde{\Omega}(\sqrt{T})$. For the problem instance, the underlying distribution $\mathcal{P}_t$ takes an infinite support. We note that when the parameter distribution has a finite support, an $O(1)$ regret bound can be derived following the approach in Vera and Banerjee (2020); Banerjee and

Freund (2020a,b), which implies that the gap between $O(\sqrt{T})$ and $O(1)$ is caused by whether the support of the parameter distribution is infinite or finite. Theorem 4.3 and Proposition 4.4 altogether state that Algorithm 3 is optimal in a worst-case sense when no additional structure is imposed on $\mathcal{P}_t$'s.

Proposition 4.4. *Under Assumption 4.1, there is no algorithm that can achieve a regret better than* $\tilde{\Omega}(\sqrt{T})$.

4.3 Non-stationary Environment with Prior Estimate: Wasserstein Based Ambiguity and Analysis

In this section, we consider a "data-driven" setting where the true distribution is unknown, but a prior estimate of the true distribution is available. However, when the prior estimate deviates from the true distribution, as is often the case in reality, then two natural questions are: (i) how to properly measure the inaccuracy of the prior estimate from the true distribution, (ii) how to design and analyze algorithm with such prior estimate. We answer these two questions in this section.

4.3.1. Wasserstein-Based Measure of Deviation. Consider the decision maker has a prior estimate/prediction $\hat{\mathcal{P}}_t$ for the true distribution $\mathcal{P}_t$ for each t, and all the predictions $\{\hat{\mathcal{P}}_t\}_{t=1}^T$ are made available at the very beginning of the procedure. We use the Wasserstein distance between $\hat{\mathcal{P}}_t$ and $\mathcal{P}_t$ to measure the deviation of the prior estimate from the true distribution. In following, we first formalize the definition and then discuss the suitability of the proposed Wasserstein-based measure.

The Wasserstein distance, also known as Kantorovich-Rubinstein metric or optimal transport distance (Villani, 2008; Galichon, 2018), is a distance function defined between probability distributions on a metric space. Its notion has a long history dating back over decades ago and gains

increasingly popularity in recent years with a wide range of applications including generative modeling (Arjovsky et al., 2017), robust optimization (Esfahani and Kuhn, 2018), statistical estimation (Blanchet et al., 2019), etc. In our context, the Wasserstein distance for two probability distributions Q_1 and Q_2 on the metric parameter space $(\Theta, \mathcal{B}_\Theta)$ is defined as follows,

$$\mathcal{W}(Q_1, Q_2) := \inf_{Q_{1,2} \in \mathcal{J}(Q_1, Q_2)} \int \rho(\theta_1, \theta_2) dQ_{1,2}(\theta_1, \theta_2) \tag{4.8}$$

where $\mathcal{J}(Q_1, Q_2)$ denotes all the joint distributions $Q_{1,2}$ for (θ_1, θ_2) that have marginals Q_1 and Q_2. The distance function $\rho(\cdot, \cdot)$ is defined earlier in (4.2).

We define the following Wasserstein-based deviation budget (WBDB) to measure the cumulative deviation of the prior estimate,

$$\mathcal{W}_T(\mathcal{P}, \hat{\mathcal{P}}) := \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \hat{\mathcal{P}}_t)$$

where $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)$ denotes the true distribution and $\hat{\mathcal{P}} = (\hat{\mathcal{P}}_1, ..., \hat{\mathcal{P}}_T)$ denotes the prior estimate.

Based on the notion of WBDB, we define a set of distributions

$$\Xi_P(W_T) := \{\mathcal{P} : \mathcal{W}_T(\mathcal{P}, \hat{\mathcal{P}}) \leqslant W_T, \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)\}$$

for a non-negative constant W_T, which we call as *deviation budget*. In this section, we consider a regret based on the set Ξ_P as defined in (4.3). In this way, the regret characterizes a worst-case performance of a certain algorithm for all the distributions $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)$ within the set Ξ_P prescribed by some W_T. Specifically, the deviation budget W_T defines the set Ξ_P by inducing an upper bound for the deviation of prior estimate. Our next theorem provides an intuitive result that W_T is an inevitable loss (in terms of the algorithm regret) as a result of the inaccuracy of the prior estimate.

Theorem 4.5. *Under Assumption 4.1, if we consider the set* $\Xi_P(W_T) := \{\mathcal{P} : \mathcal{W}_T(\mathcal{P}, \hat{\mathcal{P}}) \leqslant W_T, \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)\}$*, there is no algorithm that can achieve a regret better than* $\Omega(\max\{\sqrt{T}, W_T\})$.

Theorem 4.5 states that the lower bound of the regret is $\Omega(\max\{\sqrt{T}, W_T\})$. The theorem characterizes the best achievable algorithm performance under an inaccurate prior estimate, and precisely, the lower bound is linear in respect with the deviation of the prior estimate from the true distribution. The $\Omega(\sqrt{T})$ part inherits the result in Proposition 4.4 and it captures the intrinsic uncertainty of the underlying stochastic process over a time horizon T. The $\Omega(W_T)$ part captures the uncertainty arising from the inaccurate prior estimate.

4.3.2. Informative Gradient Descent Algorithm with Prior Estimate. Now we apply our informative gradient descent algorithm to the setting of prior estimate. A natural idea is to pretend that the prior estimate $\hat{\mathcal{P}}_t$ is indeed the true distribution $\mathcal{P}_t$. To implement the idea, we define

$$\hat{L}(\boldsymbol{p}) = \boldsymbol{c}^\top \boldsymbol{p} + \sum_{t=1}^{T} \hat{\mathcal{P}}_t h(\boldsymbol{p}; \boldsymbol{\theta})$$

where the true distribution $\mathcal{P}_t$ is replaced by its estimate $\hat{\mathcal{P}}_t$ for each component in function $L(\cdot)$. Thus it can be viewed as an approximation for the true dual function $L(\cdot)$ based on prior estimate. Let $\hat{\boldsymbol{p}}^*$ denote an optimal solution to $\hat{L}(\cdot)$,

$$\hat{\boldsymbol{p}}^* \in \operatorname{argmin}_{\boldsymbol{p} \geq 0} \hat{L}(\boldsymbol{p}) \tag{4.9}$$

and for each t, define

$$\hat{\boldsymbol{y}}_t := \hat{\mathcal{P}}_t \boldsymbol{g}(\hat{\boldsymbol{x}}(\boldsymbol{\theta}); \boldsymbol{\theta}) \;\text{ where } \hat{\boldsymbol{x}}(\boldsymbol{\theta}) = \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}} \{f(\boldsymbol{x}; \boldsymbol{\theta}) - (\hat{\boldsymbol{p}}^*)^\top \cdot \boldsymbol{g}(\boldsymbol{x}; \boldsymbol{\theta})\}. \tag{4.10}$$

Here, $\hat{\boldsymbol{y}}_t$ denotes the "optimal" expected budget consumption in the t-th time under the prior estimate. In the setting of prior estimate, we do not have the exact knowledge of the true distributions $\mathcal{P}_t$'s and therefore $\boldsymbol{y}_t$'s, so we alternatively use $\hat{\boldsymbol{y}}_t$ as a substitute. Thus, the algorithm for

the prior estimate setting, denoted by IGD($\hat{\boldsymbol{y}}$), implements Algorithm 3 with the input $\hat{\boldsymbol{y}}$ defined by (4.10). Specifically, the dual update step will become

$$\boldsymbol{p}_{t+1} = \left(\boldsymbol{p}_t + \frac{1}{\sqrt{T}}\left(\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \hat{\boldsymbol{y}}_t\right)\right) \vee \mathbf{0} \tag{4.11}$$

in Algorithm 3.

4.3.3. Regret Analysis. The analysis of IGD($\hat{\boldsymbol{y}}$) is slightly more complicated than that of IGD($\boldsymbol{y}$) in theorem 4.3 because the algorithm is built upon the function $\hat{L}(\cdot)$ defined by the prior estimate instead of the true distribution. So we first study how to bound the difference between the function $\hat{L}(\cdot)$ and $L(\cdot)$ using the deviation between prior estimate and true distribution. For a probability measure Q over the metric parameter space $(\Theta, \mathcal{B}_\Theta)$, we denote

$$L_Q(\boldsymbol{p}) \coloneqq Qh(\boldsymbol{p}; \boldsymbol{\theta}) = \int_{\boldsymbol{\theta}' \in \Theta} h(\boldsymbol{p}; \boldsymbol{\theta}') dQ(\boldsymbol{\theta}').$$

Then the function $\hat{L}(\cdot)$ and $L(\cdot)$ can be expressed as

$$\hat{L}(\boldsymbol{p}) = \boldsymbol{p}^\top \boldsymbol{c} + \sum_{t=1}^{T} L_{\hat{\mathcal{P}}_t}(\boldsymbol{p}) \quad \text{and} \quad L(\boldsymbol{p}) = \boldsymbol{p}^\top \boldsymbol{c} + \sum_{t=1}^{T} L_{\mathcal{P}_t}(\boldsymbol{p})$$

Lemma 4.5.1 states that the function $L_Q(\boldsymbol{p})$ has certain "Lipschitz continuity" in regard with the underlying distribution Q. Specifically, the supremum norm between two functions $L_{Q_1}(\boldsymbol{p})$ and $L_{Q_2}(\boldsymbol{p})$ is bounded by the Wasserstein distance between two distributions Q_1 and Q_2 up to a constant dependent on the dimension and the boundedness of the function's argument.

Lemma 4.5.1. *For two probability measures Q_1 and Q_2 over the metric parameter space $(\Theta, \mathcal{B}_\Theta)$, we have that*

$$\sup_{\boldsymbol{p} \in \Omega_{\bar{p}}} \left| L_{Q_1}(\boldsymbol{p}) - L_{Q_2}(\boldsymbol{p}) \right| \leq \max\{1, \bar{p}\} \cdot (m+1) \mathcal{W}(Q_1, Q_2) \tag{4.12}$$

where $\Omega_{\bar{p}} = [0, \bar{p}]^m$ and $\bar{p}$ is an arbitrary positive constant.

Note that the Lipschitz constant in Lemma 4.5.1 involves an upper bound of the function argument $\boldsymbol{p}$. The following lemma provides such an upper bound for the dual price $\boldsymbol{p}_t$ in IGD($\hat{\boldsymbol{\gamma}}$). The derivation is essentially the same as Lemma 4.2.1.

Lemma 4.5.2. *For each $t = 1, 2, \ldots, T$, we have that $\|\boldsymbol{p}_t\|_\infty \leq q + 1$ with probability 1, where $\boldsymbol{p}_t$ is specified by (4.11) in IGD($\hat{\boldsymbol{\gamma}}$).*

The rest of the analysis for IGD($\hat{\boldsymbol{\gamma}}$) is similar to that of IGD($\boldsymbol{\gamma}$) in Theorem 4.3. The regret of IGD($\hat{\boldsymbol{\gamma}}$) is formally stated in Theorem 4.6. Notably, the algorithm's regret matches the lower bound of Theorem 4.5 and thus it establishes the optimality of the algorithm.

Theorem 4.6. *Under Assumption 4.1, suppose a prior estimate $\hat{\mathcal{P}}$ is available and the regret is defined based on the set $\Xi_P(W_T)$, then the regret of IGD($\hat{\boldsymbol{\gamma}}$) has the following upper bound*

$$Reg_T(\pi_{IGDP}) \leqslant O(\max\{\sqrt{T}, W_T\})$$

where π_{IGDP} stands for the policy specified by IGD($\hat{\boldsymbol{\gamma}}$).

We remark the algorithm IGD($\hat{\boldsymbol{\gamma}}$) does not depend on or utilize the knowledge of the quantity W_T. On the upside, this avoids the assumption on the prior knowledge of W_T (as the knowledge of variation budget V_T (Besbes et al., 2014, 2015; Cheung et al., 2019)). On the downside, there is nothing the algorithm can do even when it knows a priori W_T is small or large. Technically, it means for our algorithm IGD($\hat{\boldsymbol{\gamma}}$), the WBNB contributes nothing in the dimension of algorithm design, and it will only influence the algorithm analysis. In particular, if we compare Theorem 4.6 with Theorem 4.3, the extra term W_T captures how the deviation of the prior estimate from the true distribution will deteriorate the performance of the gradient-based algorithm. When W_T is small, the $O(\sqrt{T})$ will be dominant and we do not need to worry about the deviation because its effect on the regret is secondary. In this light, the regret bound illuminates the effect of model misspecification/estimation error on the algorithm's performance in a non-stationary environment.

4.3.4. Sub-optimality of Static Policies. Our main algorithms – Algorithm 3 is gradient based. Compared to other existing re-solving algorithms (Jasin and Kumar, 2012; Jiang and Zhang, 2020; Arlotto and Xie, 2020; Bumpensanti and Wang, 2020; Vera and Banerjee, 2020), the gradient-based algorithms feature for simplicity and computational efficiency. In addition, the gradient-based algorithms have an adaptive and dynamic design that is crucial in stabilizing the resource consumption (i.e., not to exhaust the resource too early or have too much resource left-over). This is achieved inherently by using the realized resource consumption at each time period in the gradient update. We argue that such a dynamic design that relates the online decisions with the realized resource consumption process is necessary in achieving an optimal order of regret. In contrast, for a *static policy*, the online decisions can be dependent on the realized parameters $\boldsymbol{\theta}_t$'s but will not be affected by the dynamic of the resource consumption process. We remark that a static policy can utilize the prior estimate and be time-dependent; by "static", it means the policy remains the same regardless the realization of the resource consumption process. Examples of static policies include the bid-price policy (Talluri and Van Ryzin, 1998) and the offline-to-online policy (Cheung et al., 2020).

Definition 4.7. A static policy π is described by a set of functions $\{h_t^\pi\}_{t=1}^T$, where $h_t^\pi : \Theta \to \mathcal{X}$ is allowed to be a random function. At each period t, given the type $\boldsymbol{\theta}_t$, the policy π will take the action $h_t^\pi(\boldsymbol{\theta}_t)$ if the budget constraints are not violated.

Next, we illustrate the sub-optimality of any static policy. For the data-driven setting in Section 4.3, it is not as obvious whether a static policy can achieve the same order of regret optimality as the gradient-based algorithms. For example, what if we simply use the "offline" optimal dual solution $\boldsymbol{p}^*$ or $\hat{\boldsymbol{p}}^*$ to form a static decision rule throughout the procedure? This implements bid-price policy (Talluri and Van Ryzin, 1998) for the network revenue management problem under the known distribution setting. We can show that this bid-price policy can incur a linear regret under an environment with slight non-stationarity (arbitrarily small W_T). The following

proposition provides a more general statement on the sub-optimality of any static policy under a non-stationarity environment.

Proposition 4.8. *Suppose π is a static policy such that $\mathbb{E}_{\mathcal{H}\sim\hat{\mathcal{P}}}[R_T^{UB} - R_T(\pi)] \leq C_1 \cdot \sqrt{T}$ for some constant $C_1 > 0$, where $\hat{\mathcal{P}} = \{\hat{\mathcal{P}}_1, \ldots, \hat{\mathcal{P}}_T\}$ denotes the set of prior estimates. There exists a true distribution $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_T\}$ such that $\mathcal{W}_T(\mathcal{P}, \hat{\mathcal{P}}) \leq W_T = 4\sqrt{C_1} \cdot T^{3/4}$ and $\mathbb{E}_{\mathcal{H}\sim\mathcal{P}}[R_T^{UB} - R_T(\pi)] \geq C_2 \cdot T$ for some constant $C_2 > 0$.*

In the proposition, π denotes an arbitrary static policy that achieves an $O(\sqrt{T})$ regret when the prior estimate $\hat{\mathcal{P}}$ is accurate. But when there is a difference between the prior estimate $\hat{\mathcal{P}}$ and $\mathcal{P}$ and even if the deviation budget W_T is sublinear in T, the static policy π may still incur a linear regret for some problem instances. The implication is that for a static policy that works well under a known distribution setting, its performance can drastically deteriorate when there exists an estimation error or non-stationarity. Thus the dynamic design of gradient update or re-solving is both effective and necessary in overcoming the estimation error and the environment non-stationarity.

4.4 Non-stationary Environment Without Prior Estimate

In this section, we consider an uninformative setting where the true distribution is completely unknown to the decision maker. To one end, the discussion in this section can be viewed as a reduction of the results in the last section to a setting with an "uninformative" prior estimate. To the other, the uninformative setting draws an interesting comparison with the literature on (unconstrained) online learning/optimization in non-stationary environment (Besbes et al., 2014, 2015; Cheung et al., 2019) and its analysis exemplifies the interaction between the constraints and the non-stationarity.

4.4.1. Wasserstein-based Non-stationarity. We first illustrate how the non-stationarity over $\{\mathcal{P}_t\}_{t=1}^T$ interplays with the constraints through the following example adapted from (Golrezaei

et al., 2014). The original usage of the example in their paper is to stress the importance of balancing resource consumption in an online setting. Specifically, consider the following two linear programs as the underlying problem (PCP) for two online stochastic optimization problems,

$$\begin{aligned} \max \quad & x_1 + \ldots + x_c + (1+\kappa)x_{c+1} + \ldots + (1+\kappa)x_T && (4.13) \\ \text{s.t.} \quad & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \quad \text{for } t = 1, \ldots, T. \\ \max \quad & x_1 + \ldots + x_c + (1-\kappa)x_{c+1} + \ldots + (1-\kappa)x_T && (4.14) \\ \text{s.t.} \quad & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \quad \text{for } t = 1, \ldots, T. \end{aligned}$$

where $\kappa \in (0,1)$, $c = \frac{T}{2}$ and the true distributions for both scenarios are point-mass distributions. Without loss of generality, we assume c is an integer. For the first LP (4.13), the optimal solution is to wait and accept the later half of the orders, while for the second LP (4.14), the optimal solution is to accept the first half of the orders and deplete the resource at half time. The contrast between these two LPs (two scenarios of whether the first half or the second half is more profitable) creates difficulty for the online decision making. Without knowledge of the future orders, there is no way we can obtain a sub-linear regret in both scenarios simultaneously. Because if we exhaust too much resource in the first half of the time, then for the first scenario (4.13), we do not have enough capacity to accept all the relatively profitable orders in the second half. On the contrary, if we have too much remaining resource at the half way, then for the second scenario (4.14), those relatively profitable orders that we miss in the first half are irrevocable.

Proposition 4.9. *The worst-case regret of online constrained stochastic optimization in an adversarial setting is* $\Omega(T)$.

Proposition 4.9 states that a fully adversarial setting where $\mathcal{P}_t$ can change arbitrarily over t

does not permit a sub-linear regret. The same observation is also made in the literature (Besbes et al., 2014, 2015; Cheung et al., 2019) for unconstrained online learning problems where there is no function $\boldsymbol{g}(\cdot)$ and the decision $\boldsymbol{x}_t$ is made before the revealing of $f(\cdot)$. Specifically, Besbes et al. (2015) propose a novel measure of non-stationarity as follows (in the language of our paper),

$$V_T \coloneqq \sum_{t=1}^{T-1} TV(\mathcal{P}_t, \mathcal{P}_{t+1})$$

where $TV(\cdot,\cdot)$ denotes the total variation distance between two distributions. The quantity V_T represents the cumulative *temporal change* of the distributions by comparing $\mathcal{P}_t$ and $\mathcal{P}_{t+1}$. Unfortunately, such temporal measure fails in the constrained setting. Note that for both (4.13) and (4.14), there is only one change point throughout the whole procedure thus the non-stationarity; their *temporal change* measure is $O(1)$ but a sub-linear regret is still unattainable.

Now, we propose the definition of the Wasserstein-based non-stationarity budget (WBNB) as

$$\mathcal{W}_T(\mathcal{P}) \coloneqq \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \bar{\mathcal{P}}_T)$$

where $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)$ and $\bar{\mathcal{P}}_T$ is defined to be the uniform mixture distribution of $\{\mathcal{P}_t\}_{t=1}^T$, i.e., $\bar{\mathcal{P}}_T \coloneqq \frac{1}{T}\sum_{t=1}^T \mathcal{P}_t$. The non-stationarity measure WBNB can be viewed as a degeneration of our previous deviation measure WBDB in that WBNB replaces all the prior estimates $\hat{\mathcal{P}}_t$'s with the uniform mixture $\bar{\mathcal{P}}_T$. The caveat is that in the uninformative setting, no distribution knowledge is assumed, so the decision maker does not have access to $\bar{\mathcal{P}}_T$ unlike the prior estimate in the last section. As we will see shortly, the knowledge of $\bar{\mathcal{P}}_T$ does not affect anything in terms of the algorithm design and analysis.

Based on the notion of WBNB, we define a set of distributions

$$\Xi_U(W_T) = \{\mathcal{P} : \mathcal{W}_T(\mathcal{P}) \leqslant W_T, \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)\}$$

for a non-negative constant W_T, which we call as variation budget. Throughout this section, we consider a regret based on the set Ξ_U as defined in (4.3) in aim to characterize a "worst-case" performance of certain policy/algorithm for all the distributions in the set Ξ_U.

The variation budget W_T defines the uncertainty set Ξ_U by providing an upper bound on the non-stationarity of the distributions. Our next theorem states that it is impossible to get rid of W_T in the regret bound of any algorithm, which illustrates the sharpness of our definition of WBNB. Intuitively, it means that apart from the intrinsic stochasticity term $O(\sqrt{T})$, the (unknown) non-stationarity of the underlying distributions defined by WBNB appears to be a second bottleneck for algorithm performance. The proof of the theorem follows the same argument as Theorem 4.5.

Theorem 4.10. *Under Assumption 4.1, if we consider the set* $\Xi_U(W_T) = \{\mathcal{P} : \mathcal{W}_T(\mathcal{P}) \leqslant W_T, \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)\}$*, there is no algorithm that can achieve a regret better than* $\Omega(\max\{\sqrt{T}, W_T\})$.

4.4.2. Algorithm and Regret Analysis. One pillar of designing IGD($\boldsymbol{\gamma}$) and IGD($\hat{\boldsymbol{\gamma}}$) is the budget allocation plan $\boldsymbol{\gamma}_t$'s (or $\hat{\boldsymbol{\gamma}}_t$) prescribed by either the true distribution or the prior estimate. In the uninformative setting, the most straightforward (and probably optimal) plan is to allocate the budget evenly over the entire horizon. Algorithm 4 – *uninformative gradient descent* (UGD) – implements the intuition by evenly distributing the budget without referring to any information. Thus the UGD algorithm can be viewed as IGD$\left(\frac{c}{T}\right)$. Returning to the point mentioned earlier on the knowledge of the centric distribution $\bar{\mathcal{P}}_T$, it does not matter we know it or not; because as long as all the prior estimate distributions $\hat{\mathcal{P}}_t$ are the same over time, we always have the same budget allocation plan. Furthermore, when all the $\mathcal{P}_t$'s are the same, which means the variation budget $W_T = 0$, Algorithm 4 and its analysis collapse into several recent studies on the gradient-based online algorithm under a stationary environment (Lu et al., 2020; Li et al., 2020).

Theorem 4.11. *Under Assumption 4.1, if we consider the set* $\Xi_U(W_T) = \{\mathcal{P} : \mathcal{W}_T(\mathcal{P}) \leqslant W_T, \mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_T)\}$*, then the regret of Algorithm 4 has the following upper bound*

$$Reg_T(\pi_{UGD}) \leqslant O(\max\{\sqrt{T}, W_T\})$$

Algorithm 4 Uninformative Gradient Descent Algorithm (UGD)

1: Initialize the initial dual price $\boldsymbol{p}_1 = \mathbf{0}$ and initial constraint capacity $\boldsymbol{c}_1 = \boldsymbol{c}$.
2: **for** $t = 1, ..., T$ **do**
3: Observe $\boldsymbol{\theta}_t$ and solve

$$\tilde{\boldsymbol{x}}_t = \text{argmax}_{\boldsymbol{x}\in\mathcal{X}}\{f(\boldsymbol{x};\boldsymbol{\theta}_t) - \boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta}_t)\}$$

where $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{\theta}_t) = (g_1(\boldsymbol{x},\boldsymbol{\theta}_t), ..., g_m(\boldsymbol{x},\boldsymbol{\theta}_t))^\top$
4: Set

$$\boldsymbol{x}_t = \begin{cases} \tilde{\boldsymbol{x}}_t, & \text{if } \boldsymbol{c}_t \text{ permits a consumption of } \boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

5: Update the dual price

$$\boldsymbol{p}_{t+1} = \left(\boldsymbol{p}_t + \frac{1}{\sqrt{T}}\left(\boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - \frac{\boldsymbol{c}}{T}\right)\right) \vee \mathbf{0} \tag{4.15}$$

where the element-wise maximum operator $u \vee v = \max\{v, u\}$
6: Update the remaining capacity

$$\boldsymbol{c}_{t+1} = \boldsymbol{c}_t - \boldsymbol{g}(\boldsymbol{x}_t;\boldsymbol{\theta}_t)$$

7: **end for**
8: Output: $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$

where π_{UGD} stands for the policy specified by Algorithm 4.

Theorem 4.11 states the upper bound of Algorithm 4, which matches the regret lower bound in Theorem 4.10. Remarkably, the factors on T and W_T are additive in the regret upper bound of Algorithm 4. In comparison, the factor on T and the variation budget V_T are usually multiplicative in the regret upper bounds in the line of works that adopts the temporal change variation budget as nonstationary measure (Besbes et al., 2014, 2015; Cheung et al., 2019). The price of such an advantage for WBNB is that the WBNB is a more restrictive notion than the variation budget; recall that in (4.13) and (4.14), the temporal change variational budget is $O(1)$, but the WBNB is $O(\kappa T)$. Again, as the setting with prior estimate, the knowledge of the quantity W_T does not affect the algorithm design. By putting together Theorem 4.10 and Theorem 4.11, we argue that the knowledge of W_T does not help to further improve the algorithm performance.

5 A Primal-dual Algorithm for an Allocation Problem with Service Level Constraints

In this chapter, we develop a primal-dual algorithm for a resource allocation problem with service level constraints. In Section 5.1, we present the problem formulation. In Section 5.2, we reformulate our problem as a semi-infinite linear programming and prove its strong duality. The strong duality we proved formally shows that the *feasibility* of a capacity level is equivalent to the *asymptotic feasibility* of a capacity level, which we will discuss in detail in Section 5.2. Then, in Section 5.3, based on the duality result, we derive the optimal rationing policy for a given capacity level. We develop a minimax optimization problem to compute the optimal capacity level in Section 5.4. All the formal proofs are deferred to the appendix, however, the proof techniques are always sketched.

5.1 Problem Formulation

We consider the following general framework that models capacity allocation and demand fulfillment with *individual* service constraints. A firm serves n customers, denoted by $\mathcal{N} = \{1, 2, \cdots, n\}$. The demand of customer $j \in \mathcal{N}$ is $\tilde{D}_j$ and $\tilde{\mathbf{D}} := (\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_n)$ follows a joint distribution F with

a bounded second moment. Demand of each customer can be fulfilled by utilizing one or more types of resources from the set of m resources, denoted by $\mathcal{M} = \{1, 2, \cdots, m\}$.

The firm faces a two-stage decision problem. In the first stage, knowing the joint distribution F but not the actual demand of the customers, the firm has to decide the capacity level of the resources $\mathbf{c} := (c_1, c_2, \ldots, c_m)$, where c_i is the capacity level of resource $i \in \mathcal{M}$. The capacity investment cost is $p(\mathbf{c})$. In the second stage, the demand realizes, after which the capacity of the resources is allocated and the demand of the customers is fulfilled according to a capacity rationing policy, denoted by $\tilde{\boldsymbol{\phi}}$. We denote by $s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D})$ the fulfilled demand of customer j under policy $\tilde{\boldsymbol{\phi}}$ when the capacity level is $\boldsymbol{c}$ and the realized demand is $\boldsymbol{D} = (D_1, D_2, \ldots, D_n)$, and let $\boldsymbol{s}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}) = (s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}), j \in N)$. Notice that $s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D})$ can be a random variable even for fixed demand $\boldsymbol{D}$ if $\tilde{\boldsymbol{\phi}}$ is allowed to be a randomized policy. Similarly, we denote by $y_{ij}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D})$ the allocation of resource i to customer j and let $\mathbf{y}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}) = (y_{ij}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}), i \in \mathcal{M}, j \in \mathcal{N})$. The allocation cost is denoted by $f(\mathbf{y}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}))$. We assume that $f(\cdot)$ is a linear function throughout the paper.

Regardless of the rationing policy used, the allocation must satisfy the following constraints. More specifically, given $\boldsymbol{c}$ and $\boldsymbol{D}$, the set of all feasible fulfilled demands and resource allocation is denoted by $P(\boldsymbol{c}, \boldsymbol{D})$. By specializing the choices $P(\boldsymbol{c}, \boldsymbol{D})$, the feasible set captures the capacity consumption and demand fulfillment constraints of many capacity allocation models such as inventory pooling, process flexibility, and assemble-to-order, etc., as illustrated below.

- In inventory pooling, $\mathcal{N}$ is the set of locations, $\mathcal{M}$ is a singleton, and

$$P(\boldsymbol{c}, \boldsymbol{D}) = \left\{ (\boldsymbol{s}, \mathbf{y}) \geqslant 0 : \quad \sum_{j=1}^{n} y_j \leqslant c, \quad \mathbf{y} = \boldsymbol{s} \quad \boldsymbol{s} \leqslant \boldsymbol{D} \right\}. \tag{5.1}$$

This special case will also be referred to as the single resource allocation problem.

- In process flexibility, $\mathcal{N}$ is the set of products, $\mathcal{M}$ is the set of plants, and

$$P(\boldsymbol{c}, \boldsymbol{D}) = \left\{ (\boldsymbol{s}, \mathbf{y}) \geqslant 0 : \quad \boldsymbol{s} \leqslant \boldsymbol{D}, \sum_{j \in \mathcal{N}:(i,j) \in E} y_{ij} \leqslant c_i \ \ \forall i \in \mathcal{M}, \sum_{i \in \mathcal{M}:(i,j) \in E} y_{ij} = s_j \ \ \forall j \in \mathcal{N} \right\} \tag{5.2}$$

  where the set E represents the design of the flexible system: $(i, j) \in E$ if product j can be produced by plant i.

- In an assemble-to-order system, $\mathcal{N}$ is the set of end products, $\mathcal{M}$ is the set of components,

$$P(\boldsymbol{c}, \boldsymbol{D}) := \left\{ (\boldsymbol{s}, \mathbf{y}) \geqslant 0 : \quad A\boldsymbol{s} \leqslant \boldsymbol{c}, \quad \boldsymbol{s} \leqslant \boldsymbol{D}, \quad y_{ij} = A_{ij} s_j, \quad \forall i \in \mathcal{M}, j \in \mathcal{N} \right\} \tag{5.3}$$

  where $A_{ij} \geqslant 0$ is the amount of component i that each unit of product j requires. In a special case, the so-called generalized W-system, A is specialized as

$$A = \begin{bmatrix} I_{n \times n} \\ \mathbf{1}_n^T \end{bmatrix} \tag{5.4}$$

  where $\mathbf{1}_n$ is the n-dimensional column vector of all ones and $I_{n \times n}$ is the $n \times n$ identity matrix. In this system, each end product j requires two components, a product-specific component j and the component $n + 1$, the latter of which is common to all end products.

Clearly, not all demands can always be fulfilled, but the firm is obligated to achieve a target individual service level $\beta_j \in (0, 1)$ for each customer $j \in \mathcal{N}$. This service level constraint can be formally formulated as

$$\mathrm{E}_{\tilde{\boldsymbol{\phi}}, \tilde{\mathbf{D}}}[R_j(s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] \geqslant \beta_j \tag{5.5}$$

where $R_j : \mathcal{R}_+^2 \to \mathcal{R}_+$ is called the service measure function of customer j. Constraint (5.5) unifies different types of service level constraints in the operations management literature. For example, when $R_j(s_j, D_j) = 1_{s_j \geqslant D_j}$, constraint (5.5) defines the so-called Type I service level constraint, i.e.,

the demand of customer j must be completely satisfied with probability at least β_j. Similarly, Type II service level can be defined by letting $R_j(s_j, D_j) = s_j / \mathrm{E}[\tilde{D}_j]$, which measures the fraction of the expected demand that can be satisfied. In contrast, choosing $R_j(s_j, D_j) = s_j / D_j$ allows us to measure the fraction of the actual demand that can be fulfilled, which we name as Type III service level constraint. It is straightforward to verify that these functions all satisfy the following conditions.

Assumption 5.1.

a. For any $j \in \mathcal{N}$, $R_j(s_j, D_j)$ is non-decreasing and upper semi-continuous in s_j, for any D_j.

b. For any $\boldsymbol{c}$ and $\boldsymbol{D}$, $P(\boldsymbol{c}, \boldsymbol{D})$ is a compact set.

The firm's problem is to decide capacity level $\boldsymbol{c}$ and rationing policy $\tilde{\boldsymbol{\phi}}$ to minimize the first stage capacity investment cost $p(\mathbf{c})$ and the expected second stage allocation cost subject to the individual service constraints, which can be formulated as

$$\inf_{\boldsymbol{c} \geqslant 0, \tilde{\boldsymbol{\phi}}} \quad p(\boldsymbol{c}) + \mathrm{E}_{\tilde{\boldsymbol{\phi}}, \tilde{\mathbf{D}}}[f(\mathbf{y}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \tilde{\mathbf{D}}))] \tag{5.6}$$

$$\text{s.t.} \quad \mathrm{E}_{\tilde{\boldsymbol{\phi}}, \tilde{\mathbf{D}}}[R_j(s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] \geqslant \beta_j \quad \forall j \in \mathcal{N} \tag{5.6a}$$

$$(s_j(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D}), \mathbf{y}(\tilde{\boldsymbol{\phi}}, \boldsymbol{c}, \boldsymbol{D})) \in P(\boldsymbol{c}, \boldsymbol{D}) \quad \forall \boldsymbol{D}.$$

The formulation should also specify the set of feasible (randomized) policies, which will be discussed in the next section.

Although (5.6) is formulated as a single-period model, it is possible to approximate it by a periodic-review infinite time horizon problem as follows. Assume that the capacity is perishable, i.e., unused capacity in the previous period can not be used to satisfy future demands, and unmet demands are lost. Denote the demand in period t by $\tilde{\mathbf{D}}^{(t)}$ and assume $\tilde{\mathbf{D}}^{(1)}, \tilde{\mathbf{D}}^{(2)}, \cdots, \tilde{\mathbf{D}}^{(t)}, \cdots$, are i.i.d random variables. The fulfilled demand of customer j and the allocation in period t is

denoted by

$$s_j^{(t)}(\boldsymbol{c}, \boldsymbol{D}^{(1:t)}, \boldsymbol{s}^{(1:t-1)}, \mathbf{y}^{(1:t-1)}) \text{ and } \mathbf{y}^{(t)}(\boldsymbol{c}, \boldsymbol{D}^{(1:t)}, \boldsymbol{s}^{(1:t-1)}, \mathbf{y}^{(1:t-1)})$$

where $\boldsymbol{D}^{(1:t)} = (\boldsymbol{D}^{(1)}, \cdots, \boldsymbol{D}^{(t)})$, $\boldsymbol{s}^{(1:t-1)} = (\boldsymbol{s}^{(1)}, \cdots, \boldsymbol{s}^{(t-1)})$ and $\mathbf{y}^{(1:t-1)} = (\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(t-1)})$, which shows explicitly that the demand fulfillment and resource allocation decisions in period t can depend on realized demands up to time t and on previous fulfillment and resource allocation decisions up to time $t-1$. With these notations, formulation (5.6) can be approximated by

$$\inf_{\boldsymbol{c}\geqslant 0, \boldsymbol{s}\geqslant 0, \mathbf{y}\geqslant 0} \quad p(\boldsymbol{c}) + \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(\mathbf{y}^{(t)}(\boldsymbol{c}, \boldsymbol{D}^{(1:t)}, \boldsymbol{s}^{(1:t-1)}, \mathbf{y}^{(1:t-1)})) \tag{5.7}$$

$$\text{s.t.} \quad \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R_j(s_j^{(t)}(\boldsymbol{c}, \boldsymbol{D}^{(1:t)}, \boldsymbol{s}^{(1:t-1)}, \mathbf{y}^{(1:t-1)}), D_j^{(t)}) \geq \beta_j, \quad \forall j \in \mathcal{N} \tag{5.7a}$$

Note that the service level constraint (5.7a) is defined in an asymptotic sense, which implies the following definition of *asymptotic* feasibility for the single-period model.

Definition 5.2. A capacity level $\boldsymbol{c}$ is *asymptotically* feasible as long as for any $\epsilon > 0$, there exists a rationing policy $\tilde{\boldsymbol{\phi}}_\epsilon$ such that

$$\mathrm{E}_{\tilde{\boldsymbol{\phi}}_\epsilon, \tilde{\mathbf{D}}}[R_j(s_j(\tilde{\boldsymbol{\phi}}_\epsilon, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] \geqslant \beta_j - \epsilon, \quad \forall j \in \mathcal{N} \tag{5.8}$$

Obviously, asymptotic feasibility does not immediately imply feasibility, i.e., constraint (5.6a). However, we shall prove that under an additional yet mild assumption stated in the next section, asymptotic feasibility is equivalent to feasibility, i.e., constraint (5.8) is equivalent to constraint (5.6a). To the best of our knowledge, this is the first time that such an equivalence for the service level constraint is formally proved in the literature.

5.2 Reformulation and Strong Duality

We begin this section by formally formulating the set of feasible rationing policies in formulation (5.6). A *deterministic* policy is a function $\boldsymbol{\phi}$ from $\mathbf{R}_+^{m+n}$ to $\mathbf{R}_+^n \times \mathbf{R}_+^{nm}$ such that for any capacity level $\boldsymbol{c}$ and any realized demand $\boldsymbol{D}$,

$$\boldsymbol{\phi}(\boldsymbol{c}, \boldsymbol{D}) = (\boldsymbol{s}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D}), \mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D})) \in P(\boldsymbol{c}, \boldsymbol{D}).$$

We also denote by $\boldsymbol{s}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D})$ the demand fulfillment and $\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D})$ the allocation under the deterministic policy $\boldsymbol{\phi}$ for fixed $\boldsymbol{c}$ and $\boldsymbol{D}$. We denote the set of all deterministic policies by Φ.

A *randomized* policy is determined by a probability measure λ over Φ such that any (measurable) subset of deterministic policies $\hat{\Phi} \subseteq \Phi$ is chosen with probability $\lambda(\hat{\Phi})$. Such a randomized policy is denoted by $\tilde{\boldsymbol{\phi}}_\lambda$ or simply λ. Therefore, optimization over randomized policies can be reformulated as an optimization over probability measures. Under a deterministic policy $\boldsymbol{\phi} \in \Phi$, the service level of customer j is given by $E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)]$. Therefore, under a randomized policy $\tilde{\boldsymbol{\phi}}_\lambda$, the service level of customer j is

$$\mathrm{E}_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{\lambda}}, \tilde{\mathbf{D}}}[R_j(s_j(\tilde{\boldsymbol{\phi}}_{\boldsymbol{\lambda}}, \boldsymbol{c}, \tilde{\mathbf{D}}), D_j)] = \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi}). \tag{5.9}$$

It follows that problem (5.6) can be reformulated as

$$\inf_{\boldsymbol{c} \geqslant 0, \lambda \in \chi} \quad p(\boldsymbol{c}) + \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\lambda(\boldsymbol{\phi}) \tag{5.10}$$

$$\text{s.t.} \quad \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi}) \geqslant \beta_j \quad \forall j \in \mathcal{N} \tag{5.10a}$$

where $\chi = \{\lambda \geq 0 : \int_{\boldsymbol{\phi} \in \Phi} d\lambda(\boldsymbol{\phi}) = 1\}$. Notice that constraint (5.10a) appears to be bilinear in $(E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)], d\lambda)$, which is non-convex.

In the remainder of this section, we assume that capacity $\boldsymbol{c}$ is fixed and establish conditions for checking feasibility of a given capacity level. We also present in Section 5.3 an optimal rationing policy when the given capacity level is feasible. These results will be used in Section 5.4 for the computation of optimal or near-optimal capacity levels. The results can be obtained by considering the following semi-infinite linear program

$$\inf_{\lambda \in \chi} \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\lambda(\boldsymbol{\phi}) \tag{5.11}$$

$$\text{s.t.} \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi}) \geq \beta_j, \quad \forall j \in \mathcal{N} \tag{5.11a}$$

and its Lagrangian dual formulation. Introducing the Lagrangian dual multipliers w_j for constraints (5.11a) for each $j \in \mathcal{N}$, we obtain the Lagrangian dual formulation of (5.11)

$$\sup_{\mathbf{w} \geq 0} \inf_{\lambda \in \chi} L(\mathbf{w}, \lambda) \tag{5.12}$$

where

$$L(\mathbf{w}, \lambda) := \sum_{j \in \mathcal{N}} w_j \cdot \beta_j + \int_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi}) d\lambda(\boldsymbol{\phi}) \tag{5.13}$$

denotes the Lagrangian dual function and

$$F(\mathbf{w}, \boldsymbol{\phi}) := E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] - \sum_{j \in \mathcal{N}} w_j \cdot E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)]. \tag{5.14}$$

It is clear that for fixed $\mathbf{w}$, $\inf_{\lambda \in \chi} L(\mathbf{w}, \lambda) = \sum_{j \in \mathcal{N}} w_j \cdot \beta_j + \inf_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi})$ and thus the dual problem can be reformulated as

$$\sup_{\mathbf{w} \geq 0} \inf_{\lambda \in \chi} L(\mathbf{w}, \lambda) = \sup_{\mathbf{w} \geq 0} \sum_{j \in \mathcal{N}} w_j \cdot \beta_j + \inf_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi}) \tag{5.15}$$

Indeed, the above dual formulation can be further simplified. To that end, We first define a de-

terministic optimization problem which we call the Max-Weighted-Service problem. Specifically, for any given $\mathbf{w} \geq 0, \boldsymbol{c}$, and $\boldsymbol{D}$, define

$$g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}) = \min_{(\mathbf{y}, \boldsymbol{s}) \in P(\boldsymbol{c}, \boldsymbol{D})} f(\mathbf{y}) - \sum_{j \in \mathcal{N}} w_j R_j(s_j, D_j) \tag{5.16}$$

Under Assumption 5.1, problem (5.16) always attains its minimum in the compact set $P(\boldsymbol{c}, \boldsymbol{D})$. For any $\mathbf{w} \geq 0$, we define a deterministic policy $\boldsymbol{\phi}_{\mathbf{w}}$ such that $\boldsymbol{\phi}_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}) = (\boldsymbol{s}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}), \mathbf{y}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}))$ for any $\boldsymbol{c}$ and $\boldsymbol{D}$, where $(\boldsymbol{s}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}), \mathbf{y}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}))$ denotes an optimal solution of (5.16). (When (5.16) has multiple optimal solutions, ties are broken arbitrarily so that $(\boldsymbol{s}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}), \mathbf{y}^*_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{D}))$ is uniquely defined.)

We show in the next lemma that it suffices to focus on the Max-Weighted-Service problem to solve the dual problem (5.15).

Lemma 5.2.1. *For any fixed* $\mathbf{w} \geqslant 0$*, it holds that*

$$\inf_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi}) = F(\mathbf{w}, \boldsymbol{\phi}_{\mathbf{w}}) = E_{\tilde{\mathbf{D}}}[g(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})]. \tag{5.17}$$

In fact, when considering the feasibility of a given capacity level, we can further simplify (5.12) by assuming zero allocation cost. We are now ready to present our first result regarding the asymptotic feasibility of a given capacity level $\boldsymbol{c}$.

Theorem 5.3. *Under Assumption 5.1, a given capacity level* $\boldsymbol{c}$ *is asymptotically feasible if and only if*

$$E_{\tilde{\mathbf{D}}}[\max_{(\mathbf{y}, \boldsymbol{s}) \in P(\boldsymbol{c}, \boldsymbol{D})} \sum_{j \in \mathcal{N}} w_j R_j(s_j, D_j)] \geq \sum_{j \in \mathcal{N}} w_j \beta_j \quad \textit{for all } \mathbf{w} \geqslant 0 \tag{5.18}$$

Recall that asymptotic feasibility is weaker than feasibility. In what follows, we present sufficient conditions for feasibility. To that end, we prove strong duality between the primal-dual

pair (5.11) and (5.12) under the following assumption.

Assumption 5.4.

a. *For any $j \in \mathcal{N}$, the service measure function R_j satisfies:*

   - *$R_j(s_j, D_j)$ is non-decreasing in s_j, for any fixed D_j;*
   - *there exists a finite set of parameters $0 = a_{j,1} < a_{j,2} < \cdots < a_{j,K_j} = 1$ such that for any fixed D_j, $R_j(s_j, D_j)$ is linear in s_j when $s_j \in [a_{j,l}D_j, a_{j,l+1}D_j)$ for any $l \in \{1, 2, \ldots, K_j - 1\}$;*
   - *there exists a constant $C_1 > 0$ such that $R_j(s_j, D_j) \le C_1 \cdot \max\{1, D_j\}$ for any $\boldsymbol{D}$ and any $s_j \le D_j$.*

b. *$P(\boldsymbol{c}, \boldsymbol{D})$ is a bounded polyhedron of $(\mathbf{s}, \mathbf{y})$ defined by a set of linear inequalities on $(\mathbf{s}, \mathbf{y}, \boldsymbol{c}, \boldsymbol{D})$, including the constraints $\mathbf{s} \le \boldsymbol{D}$ and $(\mathbf{s}, \mathbf{y}) \ge 0$, and there exists a constant $C_2 > 0$ such that $\|(\mathbf{s}, \mathbf{y})\|_2^2 \le C_2 \cdot \|\boldsymbol{D}\|_2^2$ for any $\boldsymbol{D}$ and any $(\mathbf{s}, \mathbf{y}) \in P(\boldsymbol{c}, \boldsymbol{D})$.*

Assumption 5.4a requires $R_j(s_j, D_j)$ to be piece-wise linear in s_j for any fixed D_j. The break-points are defined based on the ratio s_j/D_j, which denotes the fraction of the fulfilled demand of customer j. It is satisfied by Type I, Type II, Type III service measure functions. Moreover, models such as inventory pooling, process flexibility and assemble-to-order all satisfy Assumption 5.4b.

Theorem 5.5. *Under Assumption 5.4, strong duality holds between (5.11) and (5.12) for any given capacity level $\boldsymbol{c} \geqslant 0$. Specifically, (5.11) is feasible if and only if the objective value of (5.12) is finite, and when (5.11) is feasible, the objective values of (5.11) and (5.12) are the same.*

We note that strong duality for semi-infinite linear programming under various conditions has been studied in the literature; see for example Shapiro (2001) and Martin et al. (2016). However, we have not found a simple way to verify these conditions. For example, in order to apply the strong duality results of Shapiro (2001), we have to show certain closedness or compactness properties of a topological space on the set of deterministic policies Φ. Our proof essentially shows certain

compactness of a related set. However, we choose to apply Sion's minimax theorem (Sion, 1958) to avoid introducing additional concepts required by existing conditions on the semi-infinite linear programming duality and our proof is slightly simpler.

5.3 Optimal Rationing Policy and Theoretical Bounds

In this section, we solve problem (5.11) for a fixed and feasible $\boldsymbol{c}$. To gain insights, recall the primal problem (5.11) and its dual (5.12). Strong duality proved in Theorem 5.5 under Assumption 5.4 implies the so-called complementary conditions (Shapiro, 2001), which states that for any optimal primal-dual solution pair $(d\lambda^*, w^*)$, $d\lambda^*(\boldsymbol{\phi}) > 0$ only if $\phi \in \operatorname{argmin}_{\boldsymbol{\phi}\in\Phi} F(\mathbf{w}^*, \boldsymbol{\phi})$. If we knew $\mathbf{w}^*$ in advance and if we could enumerate all policies in $\operatorname{argmin}_{\boldsymbol{\phi}\in\Phi} F(\mathbf{w}^*, \boldsymbol{\phi})$, then (5.11) becomes a finite-dimensional LP and we can then obtain the optimal randomized policy. However, it is not always possible to enumerate all polices in $\operatorname{argmin}_{\boldsymbol{\phi}\in\Phi} F(\mathbf{w}^*, \boldsymbol{\phi})$. Instead, our approach does not require precise knowledge of $\mathbf{w}^*$, nor the strong duality.

Our approach is to generate a random vector $\tilde{\mathbf{w}}$ and for fixed $\mathbf{w} = \tilde{\mathbf{w}}$ we solve problem (5.16) to obtain a deterministic policy $\boldsymbol{\phi}_{\mathbf{w}}$. This procedure gives us a randomized policy. The approach is motivated by applying the stochastic gradient descent algorithm (SGD) to solve the dual problem (5.12), which is reformulated as follows

$$\max_{\mathbf{w}\geq 0} \; G(\mathbf{w}) := \left\{ \sum_{j\in\mathcal{N}} w_j \cdot \beta_j + \min_{\phi\in\Phi} \left[\mathrm{E}_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] - \sum_{j\in\mathcal{N}} w_j \cdot \; \mathrm{E}_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] \right] \right\} \tag{5.19}$$

Specifically, SGD starts from any $\mathbf{w}^{(1)} \geq 0$ and for each $t = 1, 2, \ldots, T$, updates

$$w_j^{(t+1)} = \left[w_j^{(t)} + \gamma_T \cdot \frac{\partial \hat{G}(\mathbf{w}^{(t)}; \boldsymbol{D}^{(t)})}{\partial w_j} \right]^+ \quad \forall j \in \mathcal{N}$$

with a step size γ_T, where $\frac{\partial \hat{G}(\mathbf{w}^{(t)}; \boldsymbol{D}^{(t)})}{\partial w_j} = \beta_j - R_j\left(s_j\left(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}\right), D_j^{(t)}\right)$ and $\boldsymbol{D}^{(t)}$ is an inde-

pendent sample. Then, the expectation of $\frac{\sum_{t=1}^{T} \mathbf{w}^{(t)}}{T}$ will converge to $\mathbf{w}^*$ with an appropriate step size, e.g. $\gamma_T = \frac{1}{\sqrt{T}}$ (Hazan, 2019). Indeed, we can set $\tilde{\mathbf{w}}$ to be the uniform distribution over $\{\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \ldots, \mathbf{w}^{(T)}\}$ to derive our policy.

We present our policy in Algorithm 5. According to step 3 of Algorithm 5, our randomized

Algorithm 5 Max-Weighted-Service Policy

1: Generate demand samples $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(T)}\}$ independently from demand distribution F, where T is sufficiently large.

2: Starting from $\mathbf{w}^{(1)} = 0$, iteratively generate a random sequence $\{\mathbf{w}^{(2)}, \mathbf{w}^{(3)}, \ldots, \mathbf{w}^{(T+1)}\}$ as follows:

$$w_j^{(t+1)} = \left[w_j^{(t)} + \gamma_T \cdot \left(\beta_j - R_j \left(s_j \left(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \mathbf{D}^{(t)} \right), D_j^{(t)} \right) \right) \right]^+ \tag{5.20}$$

3: Draw a vector $\tilde{\mathbf{w}}_T$ from $\{\mathbf{w}^{(1)}, \mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(T)}\}$ uniformly at random. Given $\mathbf{w} = \tilde{\mathbf{w}}_T$, adopt the deterministic policy $\boldsymbol{\phi}_{\mathbf{w}}$.

capacity rationing policy selects a deterministic policy $\boldsymbol{\phi}_{\mathbf{w}^{(t)}}$, $t = 1, 2, 3, \cdots, T$, with probability $1/T$. We refer to this policy as the *Max-Weighted-Service policy*. By Lemma 5.2.1, under Assumption 5.1, for any $\mathbf{w}$, implementing the policy $\boldsymbol{\phi}_{\mathbf{w}}$ in Algorithm 1 only requires solving (5.16) for the given demand realization $\boldsymbol{D}$. In the following, for each $t = 1, \ldots, T$, we further denote an i.i.d. copy of $\tilde{\mathbf{D}}$ as $\tilde{\mathbf{D}}^t$, which is also independent of the samples $\{\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(T)}\}$.

Note that the Max-Weighted-Service policy requires to obtain T samples of the demand distribution. For any policy $\tilde{\phi}$, we denote by $\tilde{\phi}(T)$ if $\tilde{\phi}$ requires T samples of the demand distribution. Then, we call $\tilde{\phi}$ *asymptotically optimal* if and only if

$$\limsup_{T\to\infty} \mathrm{E}_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\tilde{\phi}(T), \boldsymbol{c}, \tilde{\mathbf{D}}))] \leq \text{Obj (5.11)} \text{ and } \liminf_{T\to\infty} \mathrm{E}_{\tilde{\mathbf{D}}}\left[R_j(s_j(\tilde{\phi}(T), \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)\right] \geqslant \beta_j, \quad \forall j \in \mathcal{N}.$$

where Obj(5.11) denotes the optimal value of (5.11). In order to prove the asymptotic optimality of the Max-Weighted-Service policy, it is sufficient to prove that the following two inequalities

hold amost surely:

$$\limsup_{T\to\infty} \frac{1}{T} \cdot \sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathbf{D}}^t}[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] \leq \text{Obj (5.11)} \tag{5.21}$$

and

$$\liminf_{T\to\infty} \frac{1}{T} \cdot \sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathbf{D}}^t}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t)\right] \geqslant \beta_j, \quad \forall j \in \mathcal{N}. \tag{5.22}$$

We are now ready to prove the asymptotic optimality of the Max-Weighted-Service policy under the following additional mild assumption.

Assumption 5.6. *The support of demand $\tilde{\mathbf{D}}$ is bounded, and there exists a constant C such that for each $j \in \mathcal{N}$, $R_j(s_j, D_j) \leq C$ for each D_j and each $s_j \leq D_j$.*

It is clear that Type I, Type II and Type III service measure functions all satisfy Assumption 5.6 when the support of demand $\tilde{\mathbf{D}}$ is bounded.

Theorem 5.7. *Under Assumption 5.1 and Assumption 5.6, if the capacity level $\boldsymbol{c}$ is feasible and the step size $\gamma_T = T^{-(\frac{1}{2}+\epsilon)}$ for some $\epsilon \in (0, 1/2)$, then the Max-Weighted-Service policy is asymptotic optimal, i.e., (5.21) and (5.22) hold almost surely.*

Theorem 5.7 shows the almost surely convergence. If we consider a weaker version of convergence, namely, convergence in expectation, a simple modification of the proof of Theorem 5.7 also shows the convergence rates in the objective value and the service level constraints. And the results hold under a weaker assumption than Assumption 5.6.

Corollary 5.8. *Suppose Assumption 5.1 holds and assume that there exists a constant C such that $E_{\tilde{\mathbf{D}}}[R_j(s_j, \tilde{D}_j)^2] \leqslant C$ for all $s_j \geqslant 0$. Then*

$$\frac{1}{T} \cdot \sum_{t=1}^{T} E_{\mathbf{w}^{(t)}, \tilde{\mathbf{D}}^t}[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] - Obj\,(5.11) \leqslant O(\gamma_T)$$

and for each $j \in \mathcal{N}$

$$\beta_j - \frac{1}{T} \cdot \sum_{t=1}^{T} E_{\mathbf{w}^{(t)}, \tilde{\mathbf{D}}^t} \left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t) \right] \leqslant O(\max\{\sqrt{\frac{1}{T}}, \sqrt{\frac{1}{T\gamma_T}}\})$$

By Corollary 5.8, we can use different step sizes to prove different convergence rates for the expected allocation cost and the expected service level. For example, by setting $\gamma_T = \frac{1}{\sqrt{T}}$, we get $O(\frac{1}{\sqrt{T}})$ convergence rate for the expected allocation cost and $O(\frac{1}{T^{1/4}})$ convergence rate for expected service levels. By setting $\gamma_T = \frac{1}{T^{1/3}}$, we get $O(\frac{1}{T^{1/3}})$ convergence rate for both the expected allocation cost and expected service levels. If there is no second stage allocation cost, then we can choose an arbitrary $\gamma_T > 0$ and get a convergence rate of $O(\frac{1}{\sqrt{T}})$ on the expected service levels.

5.4 Computing Optimal Capacity Level

Section 5.3 is focused on characterizing rationing policies for a given capacity level. In this section, we present algorithms to compute optimal capacity levels under Assumption 5.4. The development of the algorithm relies on the strong duality result in Theorem 5.5 for fixed $\boldsymbol{c}$, which then gives rise to a min-max stochastic programming formulation for the original capacity optimization problem (5.10). To present the formulation, we define for any $\mathbf{w}$ and $\boldsymbol{c}$,

$$H(\mathbf{w}, \boldsymbol{c}) = \mathrm{E}_{\tilde{\mathbf{D}}}[h(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})] \tag{5.23}$$

where for each $\boldsymbol{D}$, $h(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}) = p(\boldsymbol{c}) + \sum_j w_j \beta_j + g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D})$. Recall that $g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D})$ is defined in (5.16). From the strong duality established in Theorem 5.5, we have the following result.

Theorem 5.9. *Under Assumption 5.4, problem (5.10) is equivalent to*

$$\min_{\boldsymbol{c} \geq 0} \max_{\mathbf{w} \geq 0} \; H(\mathbf{w}, \boldsymbol{c}) \tag{5.24}$$

in the sense both problems share the same optimal capacity level.

Problem (5.24) is a minimax stochastic program, for which various optimization algorithms have been developed; see for example Nemirovski et al. (2009). However, in order to guarantee convergence to a globally optimal solution, usually convexity/concavity of the objective function is required. It is clear that for fixed $\boldsymbol{c}$ and $\boldsymbol{D}$, $g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D})$ is concave in $\mathbf{w}$. It then follows immediately that $H(\mathbf{w}, \boldsymbol{c})$ is concave in $\mathbf{w}$ for any fixed $\boldsymbol{c}$. However, convexity of $H(\mathbf{w}, \boldsymbol{c})$ in $\boldsymbol{c}$ can only be guaranteed with additional assumptions. Lemma 5.9.1 below presents one such assumption.

Lemma 5.9.1. *Under Assumption 5.4, if the following assumptions hold:*
(i). The investment cost function $p(\boldsymbol{c})$ is convex in $\boldsymbol{c}$.
(ii). For all $j \in N$, the service measure function $R_j(s_j, D_j)$ is concave in s_j for any fixed D_j
then $H(\mathbf{w}, \boldsymbol{c})$ is convex in $\boldsymbol{c}$ for any fixed $\mathbf{w} \geqslant 0$.

Obviously, the assumptions of Lemma 5.9.1 are satisfied for both Type II and Type III service constraints. Assuming that the optimal capacity level is bounded and C is a compact convex set containing the optimal capacity level as an interior point, we must have

$$\min_{\boldsymbol{c}\geq 0} \max_{\mathbf{w}\geq 0} \; H(\mathbf{w}, \boldsymbol{c}) = \min_{\boldsymbol{c}\in C} \max_{\mathbf{w}\geq 0} \; H(\mathbf{w}, \boldsymbol{c}) = \max_{\mathbf{w}\geq 0} \min_{\boldsymbol{c}\in C} \; H(\mathbf{w}, \boldsymbol{c})$$

When $H(\mathbf{w}, \boldsymbol{c})$ is convex in $\boldsymbol{c}$ and concave in $\mathbf{w}$, the maximin problem in the above strong duality relation is the dual problem of our original problem (5.10) in which $\boldsymbol{c}$ is a decision variable.

To proceed, we further make the following mild assumption about the objective function.

Assumption 5.10. *For any fixed $\varepsilon > 0$, there exists a compact convex set $\mathcal{W}$ in $\mathbb{R}^n$ such that*

$$\max_{\mathbf{w}\geq 0} \min_{\boldsymbol{c}\in C} H(\mathbf{w}, \boldsymbol{c}) - \varepsilon \leq \max_{\mathbf{w}\in \mathcal{W}} \min_{\boldsymbol{c}\in C} H(\mathbf{w}, \boldsymbol{c}) \leq \max_{\mathbf{w}\geq 0} \min_{\boldsymbol{c}\in C} H(\mathbf{w}, \boldsymbol{c}) + \varepsilon$$

It can be easily verified that this assumption in fact holds under Assumption 5.4 as long as the

optimal capacity level is finite. Since our focus in this section is to apply an existing algorithm to solve our problem, we skip the verification of this assumption.

We now apply the mirror descent stochastic approximation (SA) algorithm of Juditsky and Nemirovski (2011) to solve the minimax stochastic program (5.24). The presentation of the algorithm requires the following notations. Let $\|\cdot\|$ be a general norm defined in $\mathbb{R}^{2n}$, with $\|x\|_* = \sup_{\|v\|\leq 1} v^T x$ being its dual norm. Let $l : X = C \times \mathcal{W} \to \mathbb{R}$ be a distance-generating function. If $l(\cdot)$ is convex and continuous on X, the set $X^0 = \{x \in X : \ \exists u \in \mathbb{R}^{2n} \ \text{s.t.} \ x \in \operatorname{argmin}_{v\in X}[u^T v + l(v)]\}$ is convex. Suppose $l(\cdot)$ is continuously differentiable and strongly convex on X^0 with parameter 1 with respect to $\|\cdot\|$, i.e., $(x' - x)^T(\nabla l(x') - \nabla l(x)) \geq \|x' - x\|^2, \quad \forall x', x \in X^0$. The prox-function is defined by $V(x, z) = l(z) - [l(x) + \nabla l(x)^T(z - x)]$ and the prox mapping is defined by $\mathcal{P}_x : \mathbb{R}^{2n} \to X^0$ such that $\mathcal{P}_x(u) = \operatorname{argmin}_{z\in X}\{u^T(z - x) + V(x, z)\}$. There are many ways to choose distance generating functions: for example $l(x) = \sum_{k=1}^{2n} x_k \log(x_k)$ or $l(x) = \frac{1}{2}\|x\|_2^2$. In the following analysis, we adopt the Euclidean norm $\|\cdot\|_2$. For notational brevity, we define the tuple $\theta = [\mathbf{w}^T, \boldsymbol{c}^T]^T$ and denote $\partial h(\theta; \boldsymbol{D}) = \partial h(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D})$, which is an unbiased estimator of $\partial H(\mathbf{w}, \boldsymbol{c})$ represented by

$$\partial h(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}) = \begin{bmatrix} \partial_{\boldsymbol{c}} h(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}) \\ -\partial_{\mathbf{w}} h(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}) \end{bmatrix}.$$

Then the mirror descent SA algorithm is presented in Algorithm 6.

Algorithm 6 SA Algorithm Juditsky and Nemirovski (2011)

Input: initial point $\theta^{(1)}$, time horizon T, positive step size $\{\gamma^{(t)}\}_{t=1}^T$, and a sequence $\{\boldsymbol{D}^{(t)}\}_{t=1}^T$, which is a sequence of samples of $\tilde{\mathbf{D}}$.
Output: sequence $\{\theta^{(t)}\}_{t=1}^T$.
for $t = 1, ..., T$ **do**
  $\theta^{(t+1)} = \mathcal{P}_{\theta^{(t)}}(\gamma^{(t)} \cdot \partial h(\theta^{(t)}, \boldsymbol{D}^{(t)}))$
endfor

Proposition 1.7 of Juditsky and Nemirovski (2011) implies that Algorithm 6 converges to an

optimal capacity level when $E[\|\partial h(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})\|_2^2]$ is bounded and the step size $\{\gamma^{(t)}\}_{t=1}^{T}$ satisfies

$$\sum_{t=1}^{T} \gamma^{(t)} \to \infty \text{ and } \frac{\sum_{t=1}^{T} (\gamma^{(t)})^2}{\sum_{t=1}^{T} \gamma^{(t)}} \to 0 \quad \text{as } T \to \infty.$$

A common choice of step size is the constant step size $\gamma^{(t)} = \frac{\delta}{\sqrt{T}}$ for $t = 1, 2, \ldots, T$, where $\delta > 0$ is a parameter. One could also choose the step size $\gamma^{(t)} = \frac{\delta}{\sqrt{t}}$ for $t = 1, 2, \ldots, T$, which does not require a fixed total number of iterations in advance.

6 Online Learning Policy for Inventory Control

In this chapter, we present an online learning policy for an inventory control problem with lead time. In Section 6.1, we introduce the problem and present the formulation. In Section 6.2, we present out online learning policy and illustrate the main algorithmic idea. The regret bound is proved in Section 6.3. All the formal proofs are deferred to the appendix, however, the proof techniques are always sketched.

6.1 Problem Formulation

Consider a periodic-review lost-sales inventory system of a single product over a finite horizon of T periods. The demand at each period t is denoted by D_t, which belongs to the interval $[0, \bar{D}]$ and is assumed to be drawn independently from an unknown distribution $F(\cdot)$. At each period t, the company places an order of q_t, which will arrive after L (a positive integer) periods. We consider the case where the company may not receive exactly what it orders. To model the supply uncertainty, we introduce a supply function $s(q, z) : \mathbb{R}^2 \to \mathbb{R}$, and we assume that the company at period t receives a quantity given by $s(q_{t-L}, Z_t)$, where $Z_1, \ldots, Z_T$ are i.i.d. non-negative random variables with a common distribution function $G(\cdot)$, which is assumed to belong to the interval $[\underline{\alpha}, \bar{\alpha}]$ and is assumed to be unknown. We also denote by h the per-unit holding cost and denote

by b the per-unit lost-sale penalty cost. We have the following sequence of events happening at each period t:

1. At the beginning of period t, we observe the on-hand inventory level denoted by I_t and all the inventories in pipeline ordered from the supplier, denoted by $(x_{1,t}, x_{2,t}, \ldots, x_{L,t})$ where $x_{i,t}$ is the order quantity placed at period $t-L+i-1$ for $i = 1, \ldots, L$. The system state is $(I_t, x_{1,t}, x_{2,t}, \ldots, x_{L,t})$.

2. The inventory placed L periods ago arrives and the random variable Z_t is realized. Then, the on-hand inventory is increased to $I_t + s(x_{1,t}, Z_t)$.

3. The company places an order with amount q_t that will arrive at the beginning of period $t + L$.

4. The demand D_t is realized and is satisfied as much as possible by the on-hand inventory. We assume that unsatisfied demand is lost and unobservable.

The objective of the company is to minimize the cumulative holding and penalty costs. The system state is updated as follows:

$$I_{t+1} = (I_t + s(x_{1,t}, Z_t) - D_t)^+, x_{i,t+1} = x_{i+1,t} \ \forall 1 \leq i \leq L-1 \text{ and } x_{L,t+1} = q_t$$

A policy π for the company is specified by the order quantities $q_1^\pi, \ldots, q_T^\pi$. Since the lost demand is assumed to be unobserved, we assume that only the *censored demand* is known by the company, i.e., the company can only observe the sales quantity $\min\{I_t + s(x_{1,t}, Z_t), D_t\}$ instead of the realization of D_t, and when $I_{t+1} = 0$, the company does not know the volume of lost sales. Note that the *supply data can be censored* as well, since only $s(x_{1,t}, Z_t)$ can be observed rather than Z_t, and $s(x_{1,t}, Z_t)$ may only contain truncated information about Z_t. A policy π is *feasible* if and only if π is non-anticipative, i.e, for each t, q_t^π can only depend on the system state $(I_\tau^\pi, x_{1,\tau}^\pi, \ldots, x_{L,\tau}^\pi)$ for $\tau \leq t$ and the realized values of supply $(s_1, \ldots, s_t)$. Note that the distribution functions $F(\cdot)$ and $G(\cdot)$ are assumed to be unknown by the company and need to be learned on-the-fly. Then,

the cost incurred at period t for the policy π is denoted by

$$C_t^\pi = h \cdot (I_t + s(x_{1,t}, Z_t) - D_t)^+ + b \cdot (D_t - I_t - s(x_{1,t}, Z_t))^+$$

and the expected cumulative cost for the policy π is denoted by

$$C^\pi(T, L) = \sum_{t=1}^{T} \mathbb{E}[C_t^\pi] = \sum_{t=1}^{T} \mathbb{E}\left[h \cdot (I_t + s(x_{1,t}, Z_t) - D_t)^+ + b \cdot (D_t - I_t - s(x_{1,t}, Z_t))^+\right] \tag{6.1}$$

where T is used to indicate the dependency on the number of periods in the entire horizon and L is used to indicate the dependency on the lead time. Following this notation, the long-term average cost of the policy π is denoted by

$$C_\infty^\pi = \limsup_{T \to \infty} \frac{1}{T} \cdot C^\pi(T, L) \tag{6.2}$$

Following the standard conditions (Bu et al., 2020), we will assume that the initial inventory is $I_1 = 0$ and the initial pipeline is also 0, i.e., $x_{i,1} = 0$ for all $1 \leq i \leq L$.

6.1.1. Constant Order Policies and Notion of Regret. The optimal policy for minimizing the long-term reward in (6.2) is known to be very complex and computationally intractable due to the curse of dimensionality caused by the lead time L. Thus, heuristics have been developed to solve the problem approximately. In this section, we introduce the heuristics studied, namely the *constant order policies*, where the company places the same order in every period, regardless of the system state.

When the demand distribution and the supply function are unknown to the company, the optimal order quantity q^* for minimizing (6.2) cannot be directly computed. Our goal is to develop a feasible learning algorithm π. Using the optimal constant order policy π_{q^*} as the benchmark,

we measure the performance of the learning algorithm π using the following notion of regret:

$$\text{Regret}_T^{\pi} = C^{\pi}(T, L) - T \cdot C_{\infty}^{\pi_{q^*}} \tag{6.3}$$

An alternative way to define regret of online policy π is to measure the additive difference between $C^{\pi}(T, L)$ and $C^{\pi_{q^*}}(T, L)$. We remark that for each policy π, the alternative regret will be at the same order of the regret defined in (6.3) by noting that the gap between $C^{\pi_{q^*}}(T, L)$ and $T \cdot C_{\infty}^{\pi_{q^*}}$ can be bounded by $O(\sqrt{T})$ following standard concentration inequality for Markov chain with stationary distributions.

6.1.2. Random Supply Function. We consider the random supply function $s(q, Z)$ that takes one of the following four formulations:

1. $s(q, Z) = q \cdot Z$.
2. $s(q, Z) = \min\{q, Z\}$.
3. $s(q, Z) = qZ/(q + \alpha Z^{\rho})$ for $\rho \leq 1$ and $\alpha > 0$.
4. $s(q, Z) = (qk)/(q + Z)$, for some $k > 0$.

Formulation 6.1 and 6.1 covers the well-known random yield model and the random capacity model. Formulation 6.1 is introduced in Dada et al. (2007) to model a non-linear relationship between the order quantity and the supply, which covers an increasing concave relation of the output to the input over a wide range of parameters. Formulation 6.1 has been used in Cachon (2003); Tang and Kouvelis (2014) to study a situation where the supplier serves multiple firms and allocates the total output quantity, denoted by k, proportional to the firms' order quantities, denoted by q. Note that the firm is not able to observe the order quantities required by other firms, which is captured by the random variable Z.

The above four formulations have been studied in Feng and Shanthikumar (2018), which proves that all these four formulations are *stochasticlly linear in mid-point* (Definition 1 in Feng

and Shanthikumar (2018)). It has been shown in Bu et al. (2020) that the long-run average cost of the optimal policy converges to the long-run average cost of the optimal constant order policy as $L \to \infty$, with the gap decreases exponentially in the lead time L. This result justifies the efficiency of the constant order policy when the lead time L is large, and is also the reason why we set the optimal constant order policy as the benchmark in the definition of regret in (6.3).

Note that in formulations 6.1, 6.1, and 6.1, after observing $s(q, Z)$, the value of Z can be inferred. However, this does not hold for formulation 6.1, where the value of Z is truncated by the ordering quantity q. That is, if Z realizes to be higher than q, then the company can only observe q. This data censoring issue for supply uncertainty creates extra challenges for estimating the distribution of Z.

The following observation plays a critical role in addressing the supply data censoring issue, and it is a key step to develop our learning algorithm.

Observation 6.1. *If the random supply function takes one of the Formulation 6.1, 6.1, 6.1 and 6.1, then for any q and Z, as long as we observe the value of $s(q, Z)$, we know the value of $s(q', Z)$ for any $q' \leq q$.*

Clearly, for formulation 6.1, 6.1 and 6.1, this observation holds true by noting that the value of Z can actually be derived backward from the value of q and $s(q, Z)$. For formulation 6.1, $q = s(q, Z)$ implies $q \leq Z$. Then, for any $q' \leq q$, we must have $s(q', Z) = q'$. Also, $q > s(q, Z)$ implies $q > Z = s(q, Z)$. Then, for any $q' \leq q$, we have $s(q', Z) = \min\{q', s(q, Z)\}$. Therefore, we justify Theorem 6.1 also holds for formulation 6.1.

Remark: Note that when the supply is deterministic, i.e., $s(q, Z) = q$, it is a special case of our models. This is the classic lost-sales inventory system with positive lead times that is extensively studied in the literature (Bijvank and Vis, 2011).

6.2 Algorithm Description

In this section, we propose our learning algorithm to achieve the regret of optimal order. We being with re-formulating the long-run average cost of a constant order policy π_q. Note that in expression (6.1), the true value of D_t is unobservable due to lost sales and censored demand. However, we now show that in order to learn the optimal order quantity q^*, it is enough to focus solely on the on-hand inventory I_t and the supply $s(x_{1,t}, Z_t)$, which can be directly observed.

First, under the constant order policy π_q, the on-hand inventory is updated as follows:

$$I_{t+1}^{\pi_q} = (I_t^{\pi_q} + s(q, Z_t) - D_t)^+. \tag{6.4}$$

From queueing theory Asmussen (2008), the sequence $\{I_t^{\pi_q}\}_{t=1}^{\infty}$ converges in probability to a random variable $I_\infty^{\pi_q}$, which we refer to as the limiting inventory level under the constant order policy π_q, as long as the following condition is satisfied for the order quantity q:

$$\mathbb{E}_{Z\sim G}[s(q, Z)] < \mathbb{E}_{D\sim F}[D]. \tag{6.5}$$

If condition in (6.5) is not met, the system will explode and the on-hand inventory level will approach infinity. Therefore, we also impose the same condition for our analyses as shown in Assumption 6.2 below.

Assumption 6.2. *The company knows an upper bound of the optimal order quantity q^*, denoted by $\bar{q}$, that satisfies $\mathbb{E}[s(\bar{q}, Z)] < \mathbb{E}[D]$.*

Assumption 6.2 is very mild, because any value not satisfying the condition in (6.5) will cause the system to explode, therefore those values can be easily detected as suboptimal.

We have

$$C_\infty^{\pi_q} = h \cdot \mathbb{E}[I_\infty^{\pi_q} + s(q, Z) - D]^+ + b \cdot \mathbb{E}[D - s(q, Z) - I_\infty^{\pi_q}]^+$$

Moreover, it holds that

$$I_\infty^{\pi_q} =^d (I_\infty^{\pi_q} + s(q, Z) - D)^+$$

where $=^d$ denotes identical in distribution. By taking expectation over both sides of the following equation,

$$I_\infty^{\pi_q} + s(q, Z) - D = [I_\infty^{\pi_q} + s(q, Z) - D]^+ - [D - s(q, Z) - I_\infty^{\pi_q}]^+,$$

we have that

$$\begin{aligned}\mathbb{E}[I_\infty^{\pi_q}] + \mathbb{E}[s(q, Z)] - \mathbb{E}[D] &= \mathbb{E}[I_\infty^{\pi_q} + s(q, Z) - D]^+ - \mathbb{E}[D - s(q, Z) - I_\infty^{\pi_q}]^+ \\ &= \mathbb{E}[I_\infty^{\pi_q}] - \mathbb{E}[D - s(q, Z) - I_\infty^{\pi_q}]^+\end{aligned}$$

which implies that

$$C_\infty^{\pi_q} = h \cdot \mathbb{E}[I_\infty^{\pi_q}] + b \cdot \mathbb{E}[D] - b \cdot \mathbb{E}[s(q, Z)]. \tag{6.6}$$

Note that $C_\infty^{\pi_q}$ is unobservable, because the term $\mathbb{E}[D]$ in (6.6) is unobservable due to demand censoring. However, the term $\mathbb{E}[D]$ is independent of the order quantity q. In order to obtain the optimal order quantity q^*, it is equivalent to minimize the *pseudo-cost* defined as follows:

$$\hat{C}_\infty^{\pi_q} = h \cdot \mathbb{E}[I_\infty^{\pi_q}] - b \cdot \mathbb{E}[s(q, Z)] \tag{6.7}$$

over the set $Q = \{q : q \leqslant \bar{q}\}$. Here, the pseudo-cost $\hat{C}_\infty^{\pi_q}$ is observable. However, $\hat{C}_\infty^{\pi_q}$ is not convex in the order quantity q, in which case commonly used learning approaches such as SGD and bisection cannot be applied.

We now describe our learning algorithm for solving (6.7). Speaking at a high level, we transfer our problem into a multi-arm bandit problem by specifying $K + 1$ points uniformly over the interval $[0, \bar{q}]$ i.e., we specify a set of points $\mathcal{A} = \{a_1, \ldots, a_{K+1}\}$ such that $a_k = \frac{k-1}{K} \cdot \bar{q}$ for any $k = 1, \ldots, K + 1$. Then, our algorithm proceeds in epochs $n = 1, 2, \ldots$ by maintaining an active set $\mathcal{A}_n \subset \mathcal{A}$ for each epoch n. The key element of our algorithm is to guarantee that for each

epoch n and each point $a \in \mathcal{A}_n$, the gap between $\hat{C}_\infty^{\pi_a}$ and $\hat{C}_\infty^{\pi_{q^*}}$ is upper bounded by γ_n, where $\{\gamma_n\}_{n\geq 1}$ is a decreasing sequence to be determined later. To be specific, we let each epoch n contain $\max\{\frac{1}{\gamma_{n+1}^2} \cdot \log T, 3L\}$ number of time periods and the implementation of our algorithm at epoch n can be classified into the following three steps:

1. We implement the constant order policy $\pi_{a^{n*}}$, where a^{n*} is the largest element in the active set $\mathcal{A}_n$.

2. We use the censored demand to *simulate* the pseudo-cost of the policies π_a for each $a \in \mathcal{A}_n$ and we construct a confidence interval of $\hat{C}_\infty^{\pi_a}$ for each $a \in \mathcal{A}_n$.

3. We use the constructed confidence intervals to identify $\mathcal{A}_{n+1} \subset \mathcal{A}_n$ such that for each element $a \in \mathcal{A}_{n+1}$, the gap between $\hat{C}_\infty^{\pi_a}$ and $\hat{C}_\infty^{\pi_{a^*}}$ is upper bounded by $(h+b)\cdot\gamma_n$, where $\hat{C}_\infty^{\pi_{a^*}} = \min_{a\in\mathcal{A}} \hat{C}_\infty^{\pi_a}$.

Following the steps outlined above, as our learning algorithm proceeds and n increases, the active set $\mathcal{A}_n$ shrinkages and the optimal order quantity q^* is gradually approximated. Our algorithm is formally described in Algorithm 7. Note that the implementation of Algorithm 7 depends on a fixed constant κ_2. We provide further discussion on how to select κ_2 in a later section. By specifying the value of K and the sequence $\{\gamma_n\}_{n\geq 1}$, we are able to prove the following theorem regarding the regret upper bound of our algorithm, which is the main theorem of our paper.

Theorem 6.3. *Denote by π Algorithm 7 with input $K = \sqrt{T}$ and $\gamma_n = 2^{-n}$ for each $n \geq 1$. Suppose that the random supply function takes one of the four formulations specified in Section 6.2.2. Then, under Theorem 6.2, the regret of π has the following upper bound:*

$$Regret(\pi) \leq \kappa \cdot \kappa_2 \cdot (L + \sqrt{T}) \cdot \log T$$

where κ is a constant that is independent of L and T, and κ_2 is the constant used in Algorithm 7.

Remark: We note that Theorem 6.3 implies a regret bound of Algorithm 7 even compared to the *optimal policy*, when the lead time L is sufficiently large. To see this, we apply Theorem 1 in Bu et al. (2020) to show that $C_\infty^{\pi_{q^*}} - C_\infty^{\pi^*} \leq \kappa_3 \cdot \gamma^L$, where κ_3 and $\gamma \in (0,1)$ are constants and

Algorithm 7 Learning-based Constant Order Policy

Input: K and $\{\gamma_n\}_{n\geq 1}$.

Initialize $\mathcal{A}_1 = \mathcal{A}$, where $\mathcal{A} = \{a_1, \ldots, a_{K+1}\}$ such that $a_k = \frac{k-1}{K} \cdot \bar{D}$ for any $k = 1, \ldots, K+1$.

Set $\tau_n = \sum_{n'=1}^{n-1} \kappa_2 \cdot \max\{\frac{1}{\gamma_{n'+1}^2} \cdot \log T, 3L\} + 1$ as the start of epoch n for each $n \geq 1$, where κ_2 is a fixed constant.

for epoch $n = 1, 2, \ldots,$ **do**

  Identify a^{n*} as the largest element in the active set $\mathcal{A}_n$.

  for time period $t = \tau_n$ to $\tau_{n+1} - 1$ **do**

    Implement the constant order policy $\pi_{a^{n*}}$.

    Observe the value of the supply $s(x_{1,t}, Z_t)$ and the on-hand inventory level I_t.

  end for

  For each $a \in \mathcal{A}_n$, we construct $\tilde{C}_n^a$ as follows:

  - obtain the simulated supply $s(a, Z_t)$ under policy π_a for each $t = \tau_n + L, \ldots, \tau_{n+1} - 1$;
  - starting from $I^a_{\tau_n+L} = I_{\tau_n+L}$, for $t = \tau_n + L, \ldots, \tau_{n+1} - 1$, do the following:
    - if $I_{t+1} > 0$, then $I^a_{t+1} = (I^a_t + s(a, Z_t) + I_{t+1} - I_t - s(a^{n*}, Z_t))^+$;
    - if $I_{t+1} \leq 0$, then $I^a_{t+1} = 0$.
  - compute

$$\begin{aligned}\tilde{C}_n^a =& h \cdot \frac{1}{\tau_{n+1} - \tau_n - \kappa_2 \max\{\log T, 2L\}} \cdot \sum_{t=\tau_n+\kappa_2 \max\{\log T, 2L\}}^{\tau_{n+1}-1} I_t^a \\ & - b \cdot \frac{1}{\tau_{n+1} - \tau_n - \kappa_2 \max\{\log T, 2L\}} \cdot \sum_{t=\tau_n+\kappa_2 \max\{\log T, 2L\}}^{\tau_{n+1}-1} s(a, Z_t).\end{aligned}$$

  Denote by $\tilde{C}_n^* = \min_{a\in\mathcal{A}_n} \tilde{C}_n^a$ and identify the active set for epoch $n+1$.

$$\mathcal{A}_{n+1} = \{a \in \mathcal{A}_n : \tilde{C}_n^a \leq \tilde{C}_n^* + (h+b) \cdot \frac{\gamma_n}{2}\} \tag{6.8}$$

end for

π^* stands for the optimal policy, i.e. $\pi^* = \text{argmin}_\pi C_\infty^\pi$. Therefore, we have $C^\pi(T,L) - T \cdot C_\infty^{\pi^*} \leq \kappa \cdot \kappa_2 \cdot (L+\sqrt{T}) \cdot \log T + \kappa_3 \cdot T \cdot \gamma^L$, which implies that $C^\pi(T,L) - T \cdot C_\infty^{\pi^*} \leq O(L+\sqrt{T})$ when $L \geq O(\log T)$. This is the *first* time that a sublinear regret bound is derived for an online policy with respect to the optimal policy. As a result, our result justifies the efficiency of constant order policies for inventory control systems with large lead time, under an online learning environment.

In the literature, the most related results on regret convergence rates are derived from Huh et al. (2009), Zhang et al. (2020), Agrawal and Jia (2022) and Lyu et al. (2021), which study a special case of our problem with deterministic supply. The state-of-the-art regret convergence rate is $O(L\sqrt{T})$, derived in Agrawal and Jia (2022) for continuous demand benchmarked against the optimal base-stock heuristic policy and Lyu et al. (2021) for discrete demand benchmarked against the optimal capped base-stock heuristic policy. Our regret rate of $O(L + \sqrt{T})$ compares favorably with the existing results in this special case in terms of the dependence on L and T, and it is derived benchmarked against the optimal policy (instead of a heuristic policy) when $L \geqslant O(\log T)$.

Remark: Algorithm 7 can be carried out efficiently. Note that for each quantity $a \in \mathcal{A}$, the inventory level I_t^a is simulated at most once for each period $t = 1, \ldots, T$ and $|\mathcal{A}| = \sqrt{T}$. Therefore, the overall computation complexity of Algorithm 7 is upper bounded by $O(T^{\frac{3}{2}})$.

6.2.1. Discussion on the Simulation Step. In this section, we discuss why we could use the censored demand of the constant order policy $\pi_{a^{n*}}$ to simulate the pseudo-cost of the policies π_a for each $a \in \mathcal{A}_n$, as outlined in step 10 in Algorithm 7. Following Theorem 6.1, since a^{n*} is the largest element in the active set $\mathcal{A}_n$, after observing the value of $s(a^{n*}, Z_t)$, we know the value of $s(a, Z_t)$ for all $a \in \mathcal{A}_n$, for any $t = \tau_n + L, \ldots, \tau_{n+1} - 1$. Thus, we can use $s(a, Z_t)$ for any $t = \tau_n + L, \ldots, \tau_{n+1} - 1$ to approximate the term $\mathbb{E}[s(a, Z)]$ in the expression (6.7) for $\hat{C}_\infty^{\pi_a}$, for all $a \in \mathcal{A}_n$. Following Hoeffding's inequality, the approximation error can be bounded with a high probability (formalized in Section 6.3).

For any $a \in \mathcal{A}_n$, we can approximate $\mathbb{E}[I_\infty^{\pi_a}]$. We define a stochastic process $\{I_t^a\}_{t=\tau_n+L}^{\tau_{n+1}-1}$ revolv-

ing in the following way:

$$I^a_{\tau_n+L} = I_{\tau_n+L} \text{ and } I^a_{t+1} = (I^a_t + s(a, Z_t) - D_t)^+ \text{ for all } t = \tau_n + L, \ldots, \tau_{n+1} - 2 \tag{6.9}$$

Clearly, when the value of D_t is censored, we can not directly obtain the value of I^a_{t+1}. However, we now show that if the random supply function takes one of the four formulations specified in Section 6.2.2, we can use the on-hand inventory level I_t to derive the value of I^a_t, for any $t = \tau_n + L + 1, \ldots, \tau_{n+1} - 1$. Note that $\{I_t\}_{t=\tau_n+L}^{\tau_{n+1}-1}$ evolves in the following way:

$$I_{t+1} = (I_t - s(a^{n*}, Z_t) - D_t)^+. \tag{6.10}$$

Suppose that the value of I^a_t is known, we derive the value of I^a_{t+1} under the following two cases.
1. If $I_{t+1} > 0$, then from (6.10), we can obtain the value of D_t and we can derive the value of I^a_{t+1} directly following (6.9).
2. If $I_{t+1} \leq 0$, then we have $I_t \leq D_t - s(a^{n*}, Z_t)$. Note that $s(a, Z_\tau) \leq s(a^{n*}, Z_\tau)$ for all $\tau \leq t$, we must have $I^a_t \leq I_t$. Thus, we have

$$I^a_t \leq I_t \leq D_t - s(a^{n*}, Z_t) \leq D_t - s(a, Z_t)$$

which implies that $I^a_{t+1} = 0$.

The above two steps are formalized in the following lemma.

Lemma 6.3.1. *Suppose that the stochastic process $\{I^a_t\}_{t=\tau_n+L}^{\tau_{n+1}-1}$ is defined in (6.9) and denote by $\{I_t\}_{t=\tau_n+L}^{\tau_{n+1}-1}$ the on-hand inventory level evolving in (6.10). Then, the value of I^a_t can be computed iteratively for $t = \tau_n + L, \ldots, \tau_{n+1} - 2$ in the following way:*

- *if $I_{t+1} > 0$, then $I^a_{t+1} = (I^a_t + s(a, Z_t) + I_{t+1} - I_t - s(a^{n*}, Z_t))^+$;*
- *if $I_{t+1} \leq 0$, then $I^a_{t+1} = 0$.*

After deriving the value of $\{I_t^a\}_{t=\tau_n+L}^{\tau_{n+1}-1}$, we use this sequence to approximate $\mathbb{E}[I_\infty^{\pi_a}]$. The key is to establish the coupling between the stochastic process $\{I_t^a\}_{t=\tau_n+L}^{\tau_{n+1}-1}$ and another stochastic process, which we further explain in Section 6.3.

6.2.2. Discussion on the constant κ_2. Note that the implementation of Algorithm 7 depends on a fixed constant κ_2. We now discuss how should we select the value of κ_2.

In order for the regret bound in Theorem 6.3 to hold, a condition on the constant κ_2 would be $\kappa_2 \geq \delta(F, G, \bar{q})$, where $\delta(F, G, \bar{q})$ is a constant that depends solely on F, G and $\bar{q}$, and is indepedent of L and T. Though the value of $\delta(F, G, \bar{q})$ is unknown at the beginning since we assume the distributions F and G is unknown, we can simply set $\kappa_2 = \log T$ and the condition $\kappa_2 \geq \delta(F, G, \bar{q})$ will automatically be satisfied when T is large enough. Such an operation will only induce an additional multiplicative $\log T$ term into the final regret bound in Theorem 6.3. Another way is to spend the first $O(\sqrt{T})$ periods as a pure learning phase to learn the distributions F and G, and estimate an upper bound of $\delta(F, G, \bar{q})$, which is a constant independent of T and L. Such an operation will only induce an additional additive $O(\sqrt{T})$ term into the final regret bound in Theorem 6.3, which arises from the learning phase.

6.3 Analysis of Regret Bound

In this section, we prove the regret bound in Theorem 6.3. Our analysis can be classified into the following four steps:

1) we establish the Lipschitz continuity of the pseudo-cost $\hat{C}_\infty^{\pi_q}$ over q. As a result, instead of comparing with $\hat{C}_\infty^{\pi_{q^*}}$, we can compare with $\hat{C}_\infty^{\pi_{a^*}}$ where $a^* = \operatorname{argmin}_{a\in\mathcal{A}} \hat{C}_\infty^{\pi_a}$. We show that the additional regret term caused by this replacement of benchmark is at most $O(\sqrt{T})$.

2) we provide a bound over the gap between the actual pseudo-cost incurred at each epoch n and the long-term average $\hat{C}_\infty^{\pi_{a^{n*}}}$. The proof of the bound relies on a novel coupling argument between two stochastic process.

3) we denote by $\mathcal{E}$ the event that for each epoch n (except the last epoch), the pesudo-cost of each $a \in \mathcal{A}_n$ falls into the confidence interval $[\tilde{C}_n^a - (h+b)\cdot\frac{\gamma_n}{2}, \tilde{C}_n^a + (h+b)\cdot\frac{\gamma_n}{2}]$, i.e.,

$$\mathcal{E} = \{|\tilde{C}_n^a - \hat{C}_\infty^{\pi_a}| \leq (h+b)\cdot\frac{\gamma_n}{2},\ \forall a \in \mathcal{A}_n, \forall 1 \leq n \leq N-1\} \tag{6.11}$$

where N denotes the total number of epochs. We show that event $\mathcal{E}$ occurs with a high probability.

4) we show how a^* is approximated by the revolution of the active set $\mathcal{A}_n$ in (6.8), which leads to our final regret bound.

Following the above four steps, we decompose the regret of our policy π as follows:

$$\begin{aligned}
\text{Regret}(\pi) &= \sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (h\cdot\mathbb{E}[I_t^\pi] - b\cdot\mathbb{E}[s(a^{n*}, Z_t)] - \hat{C}_\infty^{\pi_{q^*}}) \\
&= \underbrace{\sum_{t=1}^{T}(\hat{C}_\infty^{\pi_{a^*}} - \hat{C}_\infty^{\pi_{q^*}})}_{I} + \underbrace{\sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (h\cdot\mathbb{E}[I_t^\pi] - b\cdot\mathbb{E}[s(a^{n*}, Z_t)] - \hat{C}_\infty^{\pi_{a^{n*}}})}_{II} \\
&\quad + \underbrace{\sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (\hat{C}_\infty^{\pi_{a^{n*}}} - \hat{C}_\infty^{\pi_{a^*}})}_{III}
\end{aligned} \tag{6.12}$$

We use the Lipschitz continuity established in the first step to bound the term I in (6.12). We use the high probability bound established in the second step to bound the term II in (6.12). We finally use step three and step four to bound the term III in (6.12).

6.3.1. Proof of Lipschitz Continuity. In this section, we establish the Lipschitz continuity of $\mathbb{E}[I_\infty^{\pi_q}]$ over order quantity q. We denote by $\hat{s}(\mu, Z) = s(q, Z)$ for q satisfying $\mathbb{E}[s(q, Z)] = \mu$. Our approach relies on existing result showing that if we interpret the psuedo-cost as a function over μ, then this function is a convex function, which implies Lipschitz continuity since μ belongs to a bounded region. Moreover, for the random supply function taking one of the four formulations

specified in Section 6.2.2, one can check that if we interpret μ as a function of q, then this function is Lipschitz continuous. Therefore, we prove the Lipschitz continuity of $\mathbb{E}[I_\infty^{\pi_q}]$. We summarize our result in the following lemma.

Lemma 6.3.2. *There exists a constant $\beta > 0$ such that for any $q_1, q_2 \in [0, \bar{q}]$, we have*

$$|\hat{C}_\infty^{\pi_{q_1}} - \hat{C}_\infty^{\pi_{q_2}}| \leq \beta \cdot |q_1 - q_2|$$

6.3.2. Gap Between Actual Pseudo Cost and Long-term Average Pseudo Cost. We provide the bound over the gap between the actual pseudo cost incurred during each epoch n and the pseudo-cost $\hat{C}_\infty^{\pi_{a^{n*}}}$. Our proof relies on establishing the coupling between the stochastic process $\{I_t\}_{t=\tau_n}^{\tau_{n+1}-1}$ and the stochastic process defined as follows:

$$\tilde{I}_{\tau_n}^{a^{n*}} =^d I_\infty^{\pi_{a^{n*}}} \text{ and } \tilde{I}_{t+1}^{a^{n*}} = (\tilde{I}_t^{a^{n*}} + s(a^{n*}, Z_t) - D_t)^+ \text{ for } t = \tau_n, \ldots, \tau_{n+1} - 2 \tag{6.13}$$

It is clear to see that the distribution of $\tilde{I}_t^{a^{n*}}$ is identical to the distribution of $I_\infty^{\pi_{a^{n*}}}$, for each $t = \tau_n, \ldots, \tau_{n+1} - 1$. The coupling argument is formalized in the following lemma.

Lemma 6.3.3. *Denote by N the total number of epochs and denote by $\mathcal{B}$ the event that $I_{\tau_n} \leq \kappa_1 \cdot \log T$ and $\tilde{I}_{\tau_n}^{a^{n*}} \leq \kappa_1 \cdot \log T$, where $\kappa_1 > 0$ is a fixed constant, and $\{I_{\tau_n+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}_{\tau_n+\kappa_2 \cdot \max\{\log T, 2L\}}^{a^{n*}}\}$, for every epoch $n \in [N]$, i.e.,*

$$\mathcal{B} = \{I_{\tau_n} \leq \kappa_1 \cdot \log T, \tilde{I}_{\tau_n}^{a^{n*}} \leq \kappa_1 \cdot \log T \text{ and } \{I_{\tau_n+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}_{\tau_n+\kappa_2 \cdot \max\{\log T, 2L\}}^{a^{n*}}\}, \ \forall n\}.$$

Then, we have that

$$P(\mathcal{B}) \geq 1 - \frac{3N}{T^2}.$$

From Lemma 6.3.3, we know that conditioning on the event $\mathcal{B}$, it holds that $I_t = \tilde{I}_t^{a^{n*}}$ for any $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1$ and any epoch $n \in [N]$. Moreover, note that the

distribution of $\tilde{I}_t^{a^{n*}}$ is identical to the distribution of $I_\infty^{a^{n*}}$. It holds that

$$\hat{C}_\infty^{\pi_{a^{n*}}} = h \cdot \mathbb{E}[\tilde{I}_t^{a^{n*}}] - b \cdot \mathbb{E}[s(a^{n*}, Z_t)], \quad \forall t = \tau_n, \ldots, \tau_{n+1} - 1, \ \forall n \in [N] \tag{6.14}$$

As a result, the expected value of I_t for $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1$ will be the same as the expected value of $I_\infty^{a^{n*}}$, which implies that the expected actual cost should be the same as the long-term average cost for $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1$. Thus, we can obtain an upper bound over the gap between the actual pseudo cost and the long-term average pseudo cost $\hat{C}_\infty^{\pi_{a^{n*}}}$, for each epoch $n \in [N]$. By summing up the bound for each epoch $n \in [N]$, we get an upper bound of the term II in (6.12) for the entire horizon, which is formalized in the following lemma.

Lemma 6.3.4. *It holds that*

$$\sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (h \cdot \mathbb{E}[I_t^\pi] - b \cdot \mathbb{E}[s(a^{n*}, Z_t)] - \hat{C}_\infty^{\pi_{a^{n*}}}) \leq hN \cdot \kappa_1 \kappa_2 \log T \cdot \max\{\log T, 2L\} + 3hN\bar{D}$$

6.3.3. Probability Bound on the Event $\mathcal{E}$. We now show that the pesudo-cost of each $a \in \mathcal{A}_n$ at each epoch n falls into the confidence interval $[\tilde{C}_n^a - \gamma_n, \tilde{C}_n^a + \gamma_n]$ with a high probability and we provide a bound over the probability that event $\mathcal{E}$ happens. The key is to establish the stochastic coupling between the stochastic process $\{I_t^a\}_{t=\tau_n}^{\tau_{n+1}-1}$ defined in (6.9) and the stochastic process $\{\tilde{I}_t^a\}_{t=\tau_n}^{\tau_{n+1}-1}$ defined as follows:

$$\tilde{I}_{\tau_n}^a =^d I_\infty^{\pi_a} \text{ and } \tilde{I}_{t+1}^a = (\tilde{I}_t^a + s(a, Z_t) - D_t)^+ \text{ for } t = \tau_n, \ldots, \tau_{n+1} - 2 \tag{6.15}$$

We formalize the coupling argument in the following lemma, which generalizes the stochastic coupling established in Lemma 6.3.3 from the implemented order quantity a^{n*} to all quantity $a \in \mathcal{A}_n$.

Lemma 6.3.5. *Denote by N the total number of epochs and denote by C the event that $I_{\tau_n}^a \leq \kappa_1 \cdot \log T$*

and $\tilde{I}^a_{\tau_n} \le \kappa_1 \cdot \log T$, *where* $\kappa_1 > 0$ *is a fixed constant, and* $\{I^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}} = \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}}\}$, *for every epoch* $n \in [N]$ *and every* $a \in \mathcal{A}_n$, *i.e.,*

$$C = \{I^a_{\tau_n} \le \kappa_1\cdot\log T,\ \tilde{I}^a_{\tau_n} \le \kappa_1\cdot\log T \text{ and } \{I^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}} = \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}}\},\ \forall n \in [N], \forall a \in \mathcal{A}_n\}$$

Then, we have that

$$P(C) \ge 1 - \frac{3(K+1)N}{T^2}.$$

where K *is given as the input of Algorithm 7 to denote* $|\mathcal{A}|$.

For each epoch n and each action $a \in \mathcal{A}_n$, it is clear to see that the distribution of $\tilde{I}^a_t$ is identical to the distribution of $I^{\pi_a}_\infty$. Therefore, we can use the average value of $\tilde{I}^a_t$ for $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}$ to $\tau_{n+1} - 1$ to approximate the value of $\mathbb{E}[I^{\pi_a}_\infty]$, where the length of the confidence interval can be given by γ_n. Further note that conditioning on the event C happens, the value of $\{I^a_t\}^{\tau_{n+1}-1}_{t=\tau_n+\kappa_2\cdot\max\{\log T,2L\}}$ equals the value of $\{\tilde{I}^a_t\}^{\tau_{n+1}-1}_{t=\tau_n+\kappa_2\cdot\max\{\log T,2L\}}$, which implies that $\hat{C}^{\pi_a}_\infty \in [\tilde{C}^a_n - \gamma_n, \tilde{C}^a_n + \gamma_n]$ with a high probability.

Lemma 6.3.6. *We have the following bound over the probability that event* $\mathcal{E}$ *happens, where event* $\mathcal{E}$ *is defined in* (6.11),

$$P(\mathcal{E}) \ge 1 - \frac{7(K+1)N}{T^2}.$$

6.3.4. Proof of Theorem 6.3. We are now ready to prove our main theorem. Following (6.12), we have

$$\text{Regret}(\pi) = \underbrace{\sum_{t=1}^{T}(\hat{C}^{\pi_{a^*}}_\infty - \hat{C}^{\pi_{q^*}}_\infty)}_{I} + \underbrace{\sum_n \sum_{t=\tau_n}^{\tau_{n+1}}(h\cdot\mathbb{E}[I^\pi_t] - b\cdot\mathbb{E}[s(a^{n*}, Z_t)] - \hat{C}^{\pi_{a^{n*}}}_\infty)}_{II}$$
$$+ \underbrace{\sum_n \sum_{t=\tau_n}^{\tau_{n+1}}(\hat{C}^{\pi_{a^{n*}}}_\infty - \hat{C}^{\pi_{a^*}}_\infty)}_{III}$$

We use the Lipschitz continuity established in Lemma 6.3.2 to bound the term I. We denote by $a' \in \mathcal{A}$ the nearest one to q^*. Clearly, from the construction of the set $\mathcal{A}$, we know that $|q^* - a'| \leq \frac{\bar{q}}{2K}$. Therefore, from Lemma 6.3.2, we know that

$$\mathrm{I} = T \cdot (\hat{C}_\infty^{\pi_{a^*}} - \hat{C}_\infty^{\pi_{q^*}}) \leq T \cdot (\hat{C}_\infty^{\pi_{a'}} - \hat{C}_\infty^{\pi_{q^*}}) \leq T \cdot \frac{\beta \bar{q}}{2K} = \frac{\beta \bar{q} \sqrt{T}}{2}. \tag{6.16}$$

where we note $K = \sqrt{T}$.

We now bound the term II. From Lemma 6.3.4, we know that

$$\mathrm{II} = \sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (h \cdot \mathbb{E}[I_t^\pi] - b \cdot \mathbb{E}[s(a^{n*}, Z_t)] - \hat{C}_\infty^{\pi_{a^{n*}}}) \leq hN \cdot \kappa_1 \kappa_2 \log T \cdot \max\{\log T, 2L\} + 3hN\bar{D} \tag{6.17}$$

We now proceed to bound the term III with the help of the probability bound established in Section 6.3.

We now assume that the event

$$\mathcal{E} = \{|\tilde{C}_n^a - \hat{C}_\infty^{\pi_a}| \leq (h+b) \cdot \frac{\gamma_n}{2},\ \forall a \in \mathcal{A}_n, \forall 1 \leq n \leq N-1\}$$

happens. For each epoch n and each $a \in \mathcal{A}_{n+1}$, from (6.8) and the conditions of event $\mathcal{E}$, we have

$$\hat{C}_\infty^{\pi_a} - \hat{C}_\infty^{\pi_{a^*}} \leq \tilde{C}_n^a - \tilde{C}_n^{a^*} + (h+b) \cdot \gamma_n \leq \frac{3}{2} \cdot (h+b) \cdot \gamma_n.$$

Note that $a^{(n+1)*} \in \mathcal{A}_{n+1}$, we have that

$$\hat{C}_\infty^{\pi_{a^{(n+1)*}}} - \hat{C}_\infty^{\pi_{a^*}} \leq \frac{3}{2} \cdot (h+b) \cdot \gamma_n.$$

which implies the following inequality conditional on the event $\mathcal{E}$ happens,

$$\text{III} = \sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (\hat{C}_\infty^{\pi_{a^{n*}}} - \hat{C}_\infty^{\pi_{a^*}}) \leq \frac{3(h+b)}{2} \cdot \sum_{n=1}^{N} \sum_{t=\tau_n}^{\tau_{n+1}} \gamma_{n-1}$$

Moreover, denote by N the total number of epochs. We have

$$\kappa_2 \cdot \sum_{n=1}^{N-1} \frac{1}{\gamma_n^2} \cdot \log T \leq \sum_{n=1}^{N-1} \kappa_2 \cdot \max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} \leq T$$

which implies that

$$\sum_{n=1}^{N-1} \frac{1}{\gamma_n^2} \leq \frac{T}{\kappa_2 \cdot \log T}$$

By specifying $\gamma_n = 2^{-n}$, we have that $N \leq \log_4 \frac{3T+\log T}{\kappa_2 \cdot \log T}$. Therefore, conditional on the event $\mathcal{E}$ happens, we have that

$$\begin{aligned}
\sum_{n=1}^{N} \sum_{t=\tau_n}^{\tau_{n+1}-1} \hat{C}_\infty^{a^{n*}} - \hat{C}_\infty^{a^*} &\leq \frac{3(h+b)}{2} \cdot \sum_{n=1}^{N} \sum_{t=\tau_n}^{\tau_{n+1}-1} \gamma_{n-1} = 3(h+b) \cdot \sum_{n=1}^{N} \gamma_n \cdot \max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} \qquad (6.18) \\
&= 3(h+b) \cdot \sum_{n=1}^{\lfloor \log_4 \frac{3L}{\log T} \rfloor} \gamma_n \cdot 3L + 3(h+b) \cdot \sum_{n=\lfloor \log_4 \frac{3L}{\log T} \rfloor + 1}^{N} \frac{\log T}{\gamma_n} \\
&\leq 3(h+b) \cdot \sum_{n=1}^{\lfloor \log_4 \frac{3L}{\log T} \rfloor} \gamma_n \cdot 3L + 3(h+b) \cdot \sum_{n=1}^{N} \frac{\log T}{\gamma_n} \\
&\leq 9(h+b)L + 3(h+b)(2^{N+1} - 1) \log T \\
&\leq 9(h+b)L + 6(h+b) \cdot \sqrt{\frac{(3T + \log T) \log T}{\kappa_2}}
\end{aligned}$$

If the event $\mathcal{E}$ does not happen, clearly, we have that

$$\text{III} = \sum_n \sum_{t=\tau_n}^{\tau_{n+1}} (\hat{C}_\infty^{\pi_{a^{n*}}} - \hat{C}_\infty^{\pi_{a^*}}) \leq T \cdot (h+b) \cdot \bar{D}$$

where we note that $\hat{C}_{\infty}^{\pi_{a^{n*}}} \leq (h+b)\cdot \bar{D}$ for each n. Therefore, we have the following upper bound over the term III,

$$\begin{aligned}
\text{III} &= \mathbb{E}\left[\sum_{n}\sum_{t=\tau_n}^{\tau_{n+1}}(\hat{C}_{\infty}^{\pi_{a^{n*}}} - \hat{C}_{\infty}^{\pi_{a^*}}) \mid \mathcal{E}\right]\cdot P(\mathcal{E}) + \mathbb{E}\left[\sum_{n}\sum_{t=\tau_n}^{\tau_{n+1}}(\hat{C}_{\infty}^{\pi_{a^{n*}}} - \hat{C}_{\infty}^{\pi_{a^*}}) \mid \mathcal{E}^c\right]\cdot (1-P(\mathcal{E})) \qquad (6.19)\\
&\leq \mathbb{E}\left[\sum_{n}\sum_{t=\tau_n}^{\tau_{n+1}}(\hat{C}_{\infty}^{\pi_{a^{n*}}} - \hat{C}_{\infty}^{\pi_{a^*}}) \mid \mathcal{E}\right] + T\cdot(h+b)\cdot\bar{D}\cdot(1-P(\mathcal{E}))\\
&\leq 9(h+b)L + 6(h+b)\cdot\sqrt{\frac{(3T+\log T)\log T}{\kappa_2}} + \frac{7(K+1)N\bar{D}(h+b)}{T}
\end{aligned}$$

where the last inequality follows from (6.18) and the probability bound on the event $\mathcal{E}$ from Lemma 6.3.6. Combining (6.16), (6.17), and (6.19), we have that

$$\text{Regret}(\pi) = \text{I} + \text{II} + \text{III} \leq \kappa\cdot\kappa_2\cdot(L+\sqrt{T})\cdot\log T$$

where κ is a constant that is independent of L and T. Therefore, our proof of our main result Theorem 6.3 is completed.

Bibliography

Agrawal, Shipra, Nikhil R Devanur. 2014. Fast algorithms for online stochastic convex programming. *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 1405–1424.

Agrawal, Shipra, Randy Jia. 2022. Learning in structured mdps with convex cost functions: Improved regret bounds for inventory management. *Operations Research* .

Agrawal, Shipra, Zizhuo Wang, Yinyu Ye. 2014. A dynamic near-optimal algorithm for online linear programming. *Operations Research* **62**(4) 876–890.

Alaei, Saeed. 2011. Bayesian combinatorial auctions: Expanding single buyer mechanisms to many buyers. *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 512–521.

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

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

Alaei, Saeed, MohammadTaghi Hajiaghayi, Vahid Liaghat. 2013. The online stochastic generalized assignment problem. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 11–25.

Alptekinoğlu, Aydın, Arunava Banerjee, Anand Paul, Nikhil Jain. 2013. Inventory pooling to deliver differentiated service. *Manufacturing & Service Operations Management* **15**(1) 33–44.

Arjovsky, Martin, Soumith Chintala, Léon Bottou. 2017. Wasserstein generative adversarial networks. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. 214–223.

Arlotto, Alessandro, Xinchang Xie. 2020. Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards. *Stochastic Systems* .

Arnosti, Nick, Will Ma. 2021. Tight guarantees for static threshold policies in the prophet secretary problem. *arXiv preprint arXiv:2108.12893* .

Asmussen, Søren. 2008. *Applied probability and queues*, vol. 51. Springer Science & Business Media.

Azuma, Kazuoki. 1967. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series* **19**(3) 357–367.

Banerjee, Siddhartha, Daniel Freund. 2020a. Good prophets know when the end is near. *Available at SSRN 3479189* .

Banerjee, Siddhartha, Daniel Freund. 2020b. Uniform loss algorithms for online stochastic decision-making with applications to bin packing. *SIGMETRICS* .

Bassamboo, Achal, Ramandeep S Randhawa, Jan A Van Mieghem. 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Science* **56**(8) 1285–1303.

Besbes, Omar, Yonatan Gur, Assaf Zeevi. 2014. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*. 199–207.

Besbes, Omar, Yonatan Gur, Assaf Zeevi. 2015. Non-stationary stochastic optimization. *Operations research* **63**(5) 1227–1244.

Besbes, Omar, Assaf Zeevi. 2012. Blind network revenue management. *Operations research* **60**(6) 1537–1550.

Bijvank, Marco, Iris FA Vis. 2011. Lost-sales inventory theory: A review. *European Journal of Operational Research* **215**(1) 1–13.

Blanchet, Jose, Yang Kang, Karthyek Murthy. 2019. Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability* **56**(3) 830–857.

Borel, M Émile. 1909. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **27**(1) 247–271.

Bu, Jinzhi, Xiting Gong, Dacheng Yao. 2020. Constant-order policies for lost-sales inventory models with random supply functions: Asymptotics and heuristic. *Operations Research* **68**(4) 1063–1073.

Bumpensanti, Pornpawee, He Wang. 2020. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science* .

Butcher, John Charles, Nicolette Goodwin. 2008. *Numerical methods for ordinary differential equations*, vol. 2. Wiley Online Library.

Cachon, Gérard P. 2003. Supply chain coordination with contracts. *Handbooks in operations research and management science* **11** 227–339.

Chawla, Shuchi, Nikhil Devanur, Thodoris Lykouris. 2020. Static pricing for multi-unit prophet inequalities. *arXiv preprint arXiv:2007.07990* .

Chen, Boxiao, Xiuli Chao, Hyun-Soo Ahn. 2019. Coordinating pricing and inventory replenishment with nonparametric demand learning. *Operations Research* **67**(4) 1035–1052.

Chen, Boxiao, Xiuli Chao, Cong Shi. 2021. Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Mathematics of Operations Research* **46**(2) 726–756.

Chen, Boxiao, Jiashuo Jiang, Jiawei Zhang, Zhengyuan Zhou. 2022. Learning to order for inventory systems with lost sales and uncertain supplies. *arXiv preprint arXiv:2207.04550* .

Chen, Boxiao, Cong Shi. 2019. Tailored base-surge policies in dual-sourcing inventory systems with demand learning. *Available at SSRN 3456834* .

Cheung, Wang Chi, Guodong Lyu, Chung-Piaw Teo, Hai Wang. 2020. Online planning with offline simulation. *Available at SSRN 3709882* .

Cheung, Wang Chi, David Simchi-Levi, Ruihao Zhu. 2019. Non-stationary reinforcement learning: The blessing of (more) optimism. *Available at SSRN 3397818* .

Correa, José, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, Tjark Vredeveld. 2017. Posted price mech-

anisms for a random stream of customers. *Proceedings of the 2017 ACM Conference on Economics and Computation*. 169–186.

Dada, Maqbool, Nicholas C Petruzzi, Leroy B Schwarz. 2007. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management* **9**(1) 9–32.

Dantzig, George B, Mukund N Thapa. 2006. *Linear programming 2: theory and extensions*. Springer Science & Business Media.

Devanur, Nikhil R, Kamal Jain, Balasubramanian Sivan, Christopher A Wilkens. 2019. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)* **66**(1) 7.

Dutting, Paul, Michal Feldman, Thomas Kesselheim, Brendan Lucier. 2020. Prophet inequalities made easy: Stochastic optimization by pricing nonstochastic inputs. *SIAM Journal on Computing* **49**(3) 540–582.

Ehsani, Soheil, MohammadTaghi Hajiaghayi, Thomas Kesselheim, Sahil Singla. 2018. Prophet secretary for combinatorial auctions and matroids. *Proceedings of the twenty-ninth annual acm-siam symposium on discrete algorithms*. SIAM, 700–714.

Esfahani, Peyman Mohajerin, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1-2) 115–166.

Feldman, Moran, Ola Svensson, Rico Zenklusen. 2021. Online contention resolution schemes with applications to bayesian selection problems. *SIAM Journal on Computing* **50**(2) 255–300.

Feng, Qi, J George Shanthikumar. 2018. Supply and demand functions in inventory models. *Operations Research* **66**(1) 77–91.

Galichon, Alfred. 2018. *Optimal transport methods in economics*. Princeton University Press.

Gallego, Guillermo, Garrett Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science* **40**(8) 999–1020.

Garivier, Aurélien, Eric Moulines. 2008. On upper-confidence bound policies for non-stationary bandit problems. *arXiv preprint arXiv:0805.3415* .

Golrezaei, Negin, Hamid Nazerzadeh, Paat Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Science* **60**(6) 1532–1551.

Gupta, Anupam, Marco Molinaro. 2014. How experts can solve lps online. *European Symposium on Algorithms*. Springer, 517–529.

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

Hazan, Elad. 2016. Introduction to online convex optimization. *Foundations and Trends in Optimization* **2**(3-4) 157–325.

Hazan, Elad. 2019. Introduction to online convex optimization. *arXiv preprint arXiv:1909.05207* .

Healy, Alexander D. 2008. Randomness-efficient sampling within nc1. *Computational Complexity* **17**(1) 3–37.

Hill, Theodore P, Robert P Kertz. 1982. Comparisons of stop rule and supremum expectations of iid random variables. *The Annals of Probability* 336–345.

Hou, I-H, Vivek Borkar, PR Kumar. 2009. A theory of qos for wireless. *Proc. IEEE Conf. Comput. Commun.* IEEE, 486–494.

Huh, Woonghee Tim, Ganesh Janakiraman, John A Muckstadt, Paat Rusmevichientong. 2009. An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand. *Mathematics of Operations Research* **34**(2) 397–416.

Huh, Woonghee Tim, Paat Rusmevichientong. 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1) 103–123.

Jasin, Stefanus. 2015. Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research* **63**(4) 909–915.

Jasin, Stefanus, Sunil Kumar. 2012. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* **37**(2) 313–345.

Jiang, Jiashuo, Xiaocheng Li, Jiawei Zhang. 2020. Online stochastic optimization with wasserstein based non-stationarity. *arXiv preprint arXiv:2012.06961* .

Jiang, Jiashuo, Will Ma, Jiawei Zhang. 2022a. Tight guarantees for multi-unit prophet inequalities and online stochastic knapsack. *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1221–1246.

Jiang, Jiashuo, Will Ma, Jiawei Zhang. 2022b. Tightness without counterexamples: A new approach and new results for prophet inequalities. *arXiv preprint arXiv:2205.00588* .

Jiang, Jiashuo, Shixin Wang, Jiawei Zhang. 2019. Achieving high individual service-levels without safety stock? optimal rationing policy of pooled resources. *Optimal Rationing Policy of Pooled Resources (May 2, 2019).* .

Jiang, Jiashuo, Jiawei Zhang. 2020. Online resource allocation with stochastic resource consumption. *arXiv preprint arXiv:2012.07933* .

Juditsky, Anatoli, Arkadi Nemirovski. 2011. First order methods for nonsmooth convex large-scale optimization, i: general purpose methods. *Optimization for Machine Learning* 121–148.

Kleinberg, Robert, Seth Matthew Weinberg. 2012. Matroid prophet inequalities. *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 123–136.

Krengel, Ulrich, Louis Sucheston. 1978. On semiamarts, amarts, and processes with finite value. *Probability on Banach spaces* **4** 197–266.

Lecarpentier, Erwan, Emmanuel Rachelson. 2019. Non-stationary markov decision processes, a worst-case approach using model-based reinforcement learning. *Advances in Neural Information Processing Systems*. 7216–7225.

Li, Xiaocheng, Chunlin Sun, Yinyu Ye. 2020. Simple and fast algorithm for binary integer and online linear programming. *arXiv preprint arXiv:2003.02513* .

Lim, Andrew EB, J George Shanthikumar, ZJ Max Shen. 2006. Model uncertainty, robust optimization, and learning. *Models, Methods, and Applications for Innovative Decision Making*. INFORMS, 66–94.

Lin, Meichun, Woonghee Tim Huh, Harish Krishnan, Joline Uichanco. 2022. Data-driven newsvendor problem: Performance of the sample average approximation. *Operations Research* .

Liu, Allen, Renato Paes Leme, Martin Pál, Jon Schneider, Balasubramanian Sivan. 2021. Variable decom-

position for prophet inequalities and optimal ordering. *Proceedings of the 22nd ACM Conference on Economics and Computation*. 692–692.

Liu, Shang, Jiashuo Jiang, Xiaocheng Li. 2022. Non-stationary bandits with knapsacks. *arXiv preprint arXiv:2205.12427* .

Lu, Haihao, Santiago Balseiro, Vahab Mirrokni. 2020. Dual mirror descent for online allocation problems. *arXiv preprint arXiv:2002.10421* .

Lyu, Chengyi, Huanan Zhang, Linwei Xin. 2021. Ucb-type learning algorithms for lost-sales inventory models with lead times. *Available at SSRN 3944354* .

Lyu, Guodong, Wang-Chi Cheung, Mabel C Chou, Chung-Piaw Teo, Zhichao Zheng, Yuanguang Zhong. 2019. Capacity allocation in flexible production networks: Theory and applications. *Management Science* **65**(11) 5091–5109.

Martin, Kipp, Christopher Thomas Ryan, Matt Stern. 2016. The slater conundrum: duality and pricing in infinite-dimensional optimization. *SIAM Journal on Optimization* **26**(1) 111–138.

Mieghem, Jan A Van, Nils Rudi. 2002. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing & Service Operations Management* **4**(4) 313–335.

Molinaro, Marco, Ramamoorthi Ravi. 2013. The geometry of online packing linear programs. *Mathematics of Operations Research* **39**(1) 46–59.

Nemirovski, Arkadi, Anatoli Juditsky, Guanghui Lan, Alexander Shapiro. 2009. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization* **19**(4) 1574–1609.

Puterman, Martin L. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Russac, Yoan, Claire Vernade, Olivier Cappé. 2019. Weighted linear bandits for non-stationary environments. *Advances in Neural Information Processing Systems*. 12040–12049.

Samuel-Cahn, Ester. 1984. Comparison of threshold stop rules and maximum for independent nonnegative random variables. *the Annals of Probability* 1213–1216.

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shapiro, Alexander. 2009. Semi-infinite programming, duality, discretization and optimality conditions. *Optimization* **58**(2) 133–161.

Sion, Maurice. 1958. On general minimax theorems. *Pacific Journal of mathematics* **8**(1) 171–176.

Stein, Clifford, Van-Anh Truong, Xinshang Wang. 2020. Advance service reservations with heterogeneous customers. *Management Science* **66**(7) 2929–2950.

Sun, Rui, Xinshang Wang, Zijie Zhou. 2020. Near-optimal primal-dual algorithms for quantity-based network revenue management. *arXiv preprint arXiv:2011.06327* .

Swaminathan, Jayashankar M, Ramesh Srinivasan. 1999. Managing individual customer service constraints under stochastic demand. *Operations Research Letters* **24**(3) 115–125.

Talluri, Kalyan, Garrett Van Ryzin. 1998. An analysis of bid-price controls for network revenue management. *Management science* **44**(11-part-1) 1577–1593.

Tang, Sammi Y, Panos Kouvelis. 2014. Pay-back-revenue-sharing contract in coordinating supply chains with random yield. *Production and Operations Management* **23**(12) 2089–2102.

Topkis, Donald M. 2011. *Supermodularity and complementarity*. Princeton university press.

Vera, Alberto, Siddhartha Banerjee. 2020. The bayesian prophet: A low-regret framework for online decision making. *Management Science* .

Villani, Cédric. 2008. *Optimal transport: old and new*, vol. 338. Springer Science & Business Media.

Wang, Xinshang, Van-Anh Truong, David Bank. 2018. Online advance admission scheduling for services with customer preferences. *arXiv preprint arXiv:1805.10412* .

Yan, Qiqi. 2011. Mechanism design via correlation gap. *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 710–719.

Yuan, Hao, Qi Luo, Cong Shi. 2021. Marrying stochastic gradient descent with bandits: Learning algorithms for inventory systems with fixed costs. *Management Science* **67**(10) 6089–6115.

Zhang, Huanan, Xiuli Chao, Cong Shi. 2018. Perishable inventory systems: Convexity results for base-stock policies and learning algorithms under censored demand. *Operations Research* **66**(5) 1276–1286.

Zhang, Huanan, Xiuli Chao, Cong Shi. 2020. Closing the gap: A learning algorithm for lost-sales inventory systems with lead times. *Management Science* **66**(5) 1962–1980.

Zhong, Yuanguang, Zhichao Zheng, Mabel C Chou, Chung-Piaw Teo. 2017. Resource pooling and allocation policies to deliver differentiated service. *Management Science* **64**(4) 1555–1573.

A | Appendix for Chapter 2

Proof of Lemma 2.0.2:. The only part that requires a proof is to establish the strong duality between Primal($\boldsymbol{p}, \boldsymbol{D}$) and Dual($\boldsymbol{p}, \boldsymbol{D}$). We first show that Primal($\boldsymbol{p}, \boldsymbol{D}$) can be re-formulated as the following LP:

$$\begin{aligned}
\text{LP}(\boldsymbol{p}, \boldsymbol{D}) = \min \quad & V_1(1) && \text{(A.1)}\\
\text{s.t.} \quad & V_t(c_t) \geq V_{t+1}(c_t) + \sum_{d_t : d_t \leq c_t} p_t(d_t) \cdot W_t(d_t, c_t), \quad \forall t, \forall c_t && \text{(A.1a)}\\
& W_t(d_t, c_t) \geq r(d_t) + V_{t+1}(c_t - d_t) - V_{t+1}(c_t), \quad \forall t, \forall c_t \geq d_t && \text{(A.1b)}\\
& \sum_{t=1}^{T} \sum_{d_t} p_t(d_t) \cdot r(d_t) = 1 && \text{(A.1c)}\\
& V_{T+1}(c) = 0, \quad \forall c\\
& V_t(c_t), W_t(c_t) \geq 0, r(d_t) \geq 0 \quad \forall t, \forall c_t
\end{aligned}$$

For each t and d_t, we denote by

$$r(d_t) = \frac{1}{p_t(d_t)} \cdot \sum_{\omega_t : d(\omega_t) = d_t} r(\omega_t) \cdot p_t(\omega_t)$$

We have that

$$\sum_{t=1}^{T} \sum_{d_t} p_t(d_t) \cdot r(d_t) = \sum_{t=1}^{T} \sum_{d_t} \sum_{\omega_t : d(\omega_t) = d_t} r(\omega_t) \cdot p_t(\omega_t) = \sum_{t=1}^{T} \sum_{\omega_t} r(\omega_t) \cdot p_t(\omega_t) = 1$$

Thus, we get constraint (A.1c) in LP($\boldsymbol{p}, \boldsymbol{D}$) in (A.1). We also denote by

$$W_t(d_t, c_t) = \frac{1}{p_t(d_t)} \cdot \sum_{\omega_t : d(\omega_t) = d_t} p_t(\omega_t) \cdot W_t(\omega_t, c_t) \tag{A.2}$$

Then we have that

$$\mathbb{E}_{\omega_t \sim F_t} \left[1_{\{d(\omega_t) \le c_t\}} \cdot W_t(\omega_t, c_t) \right] = \sum_{d_t : d_t \le c_t} p_t(d_t) \cdot W_t(d_t, c_t)$$

which enables us to get constraint (A.1a) in LP($\boldsymbol{p}, \boldsymbol{D}$). Finally, multiplying both sides of (**??**) by $p(\omega_t)$ and summing over all ω_t satisfying $d(\omega_t) = d_t$, we get

$$\begin{aligned} p_t(d_t) \cdot W_t(d_t, c_t) &= \sum_{\omega_t : d(\omega_t) = d_t} p(\omega_t) \cdot W_t(\omega_t, c_t) \\ &\ge \sum_{\omega_t : d(\omega_t) = d_t} p(\omega_t) \cdot (r(\omega_t) + V_{t+1}(c_t - d(\omega_t)) - V_{t+1}(c_t)) \\ &= p_t(d_t) \cdot (r(d_t) + V_{t+1}(c_t - d_t) - V_{t+1}(c_t)) \end{aligned}$$

Thus, we get constraint (A.1b) in LP($\boldsymbol{p}, \boldsymbol{D}$). Note that the constraint (**??**) in Primal($\boldsymbol{p}, \boldsymbol{D}$) can also be recovered from constraint (A.1b) in LP($\boldsymbol{p}, \boldsymbol{D}$), by setting

$$W_t(\omega_t, c_t) = W_t(d_t, c_t) \text{ and } r(\omega_t) = r(d_t)$$

for all ω_t satisfying $d(\omega_t) = d_t$, and all d_t. Thus, we conclude that LP($\boldsymbol{p}, \boldsymbol{D}$) is a re-formulation of Primal($\boldsymbol{p}, \boldsymbol{D}$) and it holds that

$$\inf_{\hat{\mathcal{H}}} \frac{V_1^*(1)}{\mathbf{UP}(\hat{\mathcal{H}})} = \inf_{(\boldsymbol{p}, \boldsymbol{D}) : \sum_{t=1}^T \sum_{d_t} p_t(d_t) \cdot d_t \le 1} \text{Primal}(\boldsymbol{p}, \boldsymbol{D}) = \inf_{(\boldsymbol{p}, \boldsymbol{D}) : \sum_{t=1}^T \sum_{d_t} p_t(d_t) \cdot d_t \le 1} \text{LP}(\boldsymbol{p}, \boldsymbol{D})$$

It remains to show that LP($\boldsymbol{p}, \boldsymbol{D}$) = Dual($\boldsymbol{p}, \boldsymbol{D}$) for each ($\boldsymbol{p}, \boldsymbol{D}$) satisfying $\sum_{t=1}^T \sum_{d_t} p_t(d_t) \cdot d_t \le 1$.

We introduce the dual variable $\kappa_t(c_t)$ for constraint (A.1a), the dual variable $\alpha_t(d_t, c_t)$ for

constraint (A.1b) and a dual variable θ for constraint (A.1c). Then we have the following LP as the dual of LP($\boldsymbol{p}, \boldsymbol{D}$).

$$
\begin{aligned}
&\max \quad \theta && (A.3)\\
&\text{s.t.} \quad \theta \cdot p_t(d_t) \le \sum_{c_t \ge d_t} \alpha_t(d_t, c_t), \quad \forall t, \forall d_t \\
&\qquad \alpha_t(d_t, c_t) \le p_t(d_t) \cdot \kappa_t(c_t), \quad \forall t, \forall d_t, \forall c_t \ge d_t \\
&\qquad \kappa_t(c_t) = \kappa_{t-1}(c_t) + \sum_{d_{t-1}} (\alpha_{t-1}(d_{t-1}, c_t + d_{t-1}) - \alpha_{t-1}(d_{t-1}, c_t)), \quad \forall t, \forall c_t \le 1 \\
&\qquad \kappa_1(1) = 1 \\
&\qquad \alpha_t(d_t, c_t) = 0, \quad \forall t, \forall d_t, \forall c_t > 1 \\
&\qquad \theta, \alpha_t(c_t) \ge 0, \kappa_t(c_t) \ge 0, \quad \forall t, \forall c_t
\end{aligned}
$$

Note that from the binding constraints, it must hold

$$\kappa_t(c_t) = \kappa_1(c_t) + \sum_{\tau \le t-1} \sum_{d_\tau} (\alpha_\tau(d_\tau, c_t + d_\tau) - \alpha_\tau(d_\tau, c_t))$$

where we set $\kappa_1(c) = 0$ for all $c < 1$. Thus, we further simplify the LP (A.3) by eliminating variable $\kappa_t(c_t)$, and derive LP Dual($\boldsymbol{p}, \boldsymbol{D}$) in (2.4).

In order to establish the strong duality between LP($\boldsymbol{p}, \boldsymbol{D}$) and Dual($\boldsymbol{p}, \boldsymbol{D}$), we only need to check the so-called Slater's condition such that all the inequality constraints in LP($\boldsymbol{p}, \boldsymbol{D}$) can be satisfied as strict inequality. Clearly, this can be done by setting $\theta = 0$ and iteratively setting $\alpha_t(d_t, c_t)$ to be strictly smaller than the RHS of the constraints (A.1b), (A.1c) for $t = 1, \ldots, T$. Thus, our proof is completed. □

Proof of Lemma 2.3.1

We first present the following lemma, which show that instead of checking whether all the constraints of Dual($\boldsymbol{p}, \boldsymbol{D}$) are satisfied, it is enough to consider only one constraint.

Lemma A.0.1. *For any $\theta \in [0,1]$, $\{\theta, x_{l,t}(\theta)\}$ is a feasible solution to Dual($\boldsymbol{p}, k$) if and only if $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta) \leq 1-\theta$.*

The proof is relegated to Appendix A. We now prove the condition on θ such that $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta) \leq 1-\theta$. Due to Lemma A.0.1, this condition implies the feasibility condition of $\{\theta, x_{l,t}(\theta)\}$. Specifically, we will first show that the term $\sum_{t=1}^{T-1} x_{k,t}(\theta)$ is continuously monotone increasing with θ in the next lemma, where the formal proof is in **??**.

Lemma A.0.2. *For any $1 \leq l \leq k$ and any $1 \leq t \leq T$, define $y_{l,t}(\theta) = \sum_{\tau=1}^{t} x_{l,t}(\theta)$. Then $y_{l,t}(\theta)$ is monotone increasing with θ and $y_{l,t}(\theta)$ is also Lipschitz continuous with θ.*

We are now ready to prove Lemma 2.3.1.

Proof of Lemma 2.3.1:. Note that when $\theta = 0$, $y_{k,T-1}(0) = \sum_{\tau=1}^{T-1} x_{k,\tau}(0) = 0 < 1-\theta = 1$, and when $\theta = 1$, $y_{k,T-1}(1) = \sum_{\tau=1}^{T-1} x_{k,\tau}(1) > 0 = 1-\theta$. Further note that $1-\theta$ is continuously strictly decreasing with θ while Lemma A.0.2 shows that $y_{k,T-1}(\theta) = \sum_{\tau=1}^{T-1} x_{k,\tau}(\theta)$ is continuously increasing with θ, there must exist a unique $\theta^* \in [0,1]$ such that $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta^*) = 1-\theta^*$ and for any $\theta \in [0, \theta^*]$, it holds that $\sum_{\tau=1}^{T-1} x_{k,\tau}(\theta) \leq 1-\theta$. Combining the above arguments with Lemma A.0.1, we complete our proof. □

Proof of Lemma A.0.1

We first prove that for any $\theta \in [0,1]$, $\{x_{l,t}(\theta)\}$ are non-negative.

Lemma A.0.3. *For any $\theta \in [0,1]$, we have $x_{l,t}(\theta) \geq 0$ for any $l = 1, \ldots, k$ and $t = 1, \ldots, T$.*

Proof:. We now use induction on l to show that for any l, we have that $x_{l,t}(\theta) \geq 0$ and $\sum_{v=1}^{l} x_{v,t}(\theta) \leq \theta \cdot p_t$ for any t. Since we focus on a fixed θ, we abbreviate θ in the expression $x_{l,t}(\theta)$ and substitute $x_{l,t}$ for $x_{l,t}(\theta)$ in the proof.

For $l = 1$, from definition, we have that for $1 \leq t \leq t_2$, it holds that $x_{1,t} \geq 0$ and $\sum_{v=1}^{1} x_{v,t} \leq \theta \cdot p_t$. We now use induction on t to show that for $t_2 + 1 \leq t \leq T$, we have that $0 \leq x_{1,t} \leq \theta \cdot p_t$. Note that from definition, we have that

$$x_{1,t_2+1} = p_{t_2+1} \cdot (1 - \sum_{\tau=1}^{t_2} \theta \cdot p_\tau) < p_{t_2+1} \cdot \theta$$

Also, note that $1 - \sum_{\tau=1}^{t_2-1} \theta \cdot p_\tau \geq \theta$ and $p_{t_2} \leq 1$, we have that

$$1 - \sum_{\tau=1}^{t_2} \theta \cdot p_\tau \geq 1 - \sum_{\tau=1}^{t_2-1} \theta \cdot p_\tau - \theta \geq 0$$

Thus, it holds that

$$0 \leq x_{1,t_2+1} = p_{t_2+1} \cdot (1 - \sum_{\tau=1}^{t_2} \theta \cdot p_\tau) < p_{t_2+1} \cdot \theta$$

Now, suppose for a t such that $t_2 + 1 \leq t \leq T$, we have that $0 \leq x_{1,\tau} \leq \theta \cdot p_\tau$ for any $t_2 + 1 \leq \tau \leq t$. Then we have that

$$x_{1,t+1} \leq p_{t+1} \cdot (1 - \sum_{\tau=1}^{t} x_{1,\tau}) \leq p_{t+1} \cdot (1 - \sum_{\tau=1}^{t_2} x_{1,\tau}) = p_{t+1} \cdot (1 - \sum_{\tau=1}^{t_2} \theta \cdot p_\tau) < p_{t+1} \cdot \theta$$

Also, note that $x_{1,t} \geq 0$ implies that $1 - \sum_{\tau=1}^{t-1} x_{1,\tau} \geq 0$, we have that

$$x_{1,t+1}/p_{t+1} = 1 - \sum_{\tau=1}^{t-1} x_{1,\tau} - x_{1,t} = (1 - p_t) \cdot (1 - \sum_{\tau=1}^{t-1} x_{1,\tau}) \geq 0$$

It holds that $0 \leq x_{1,t+1} \leq p_{t+1} \cdot \theta$. Thus, from induction, for any t, we have proved that $0 \leq x_{1,t} \leq p_t \cdot \theta$.

Suppose that for a l such that $1 \leq l \leq k$, we have that $x_{l,t} \geq 0$ and $\sum_{v=1}^{l} x_{v,t} \leq \theta \cdot p_t$ for any t. We now consider the case for $l+1$. From definition, $x_{l+1,t} = 0$ when $1 \leq t \leq t_{l+1}$ and when $t_{l+1}+1 \leq t \leq t_{l+2}$, $x_{l+1,t} = \theta \cdot p_t - \sum_{v=1}^{l} x_{v,t}$. Thus, for $1 \leq t \leq t_{l+2}$, we have proved that $x_{l+1,t} \geq 0$ and $\sum_{v=1}^{l+1} x_{v,t} \leq \theta \cdot p_t$. We now use induction on t for $t > t_{l+2}$. When $t = t_{l+2}+1$, from definition, we have that

$$x_{l+1,t_{l+2}+1} = p_{t_{l+2}+1} \cdot \sum_{\tau=1}^{t_{l+2}} (x_{l,\tau} - x_{l+1,\tau}) \leq \theta \cdot p_{t_{l+2}+1} - \sum_{v=1}^{l} x_{v,t_{l+2}+1} \Rightarrow \sum_{v=1}^{l+1} x_{v,t_{l+2}+1} \leq \theta \cdot p_{t_{l+2}+1}$$

Also, note that

$$0 \leq x_{l+1,t_{l+2}} = \theta \cdot p_{t_{l+2}} - \sum_{v=1}^{l} x_{v,t_{l+2}} \leq p_{t_{l+2}} \cdot \sum_{\tau=1}^{t_{l+2}-1} (x_{l,\tau} - x_{l+1,\tau})$$

we get that

$$x_{l+1,t_{l+2}+1}/p_{t_{l+2}+1} = \sum_{\tau=1}^{t_{l+2}} (x_{l,\tau} - x_{l+1,\tau}) \geq \sum_{\tau=1}^{t_{l+2}-1} (x_{l,\tau} - x_{l+1,\tau}) - x_{l+1,t_{l+2}} \geq (1 - p_{t_{l+2}}) \cdot \sum_{\tau=1}^{t_{l+2}-1} (x_{l,\tau} - x_{l+1,\tau})$$

Thus, we proved that $0 \leq x_{l+1,t_{l+2}+1}$ and $\sum_{v=1}^{l+1} x_{v,t_{l+2}+1} \leq \theta \cdot p_{t_{l+2}+1}$. Now suppose that for a t such that $t_{l+2}+1 \leq t \leq T$, it holds that $0 \leq x_{l+1,t}$ and $\sum_{v=1}^{l+1} x_{v,t} \leq \theta \cdot p_t$. Then we have that

$$\sum_{v=1}^{l+1} x_{v,t+1}/p_{t+1} = 1 - \sum_{\tau=1}^{t} x_{l+1,\tau} \leq 1 - \sum_{\tau=1}^{t-1} x_{l+1,\tau} = \sum_{v=1}^{l+1} x_{v,t}/p_t \leq \theta$$

Also, note that $0 \leq x_{l+1,t} = p_t \cdot \sum_{\tau=1}^{t-1} (x_{l,\tau} - x_{l+1,\tau})$, we have that

$$x_{l+1,t+1}/p_{t+1} = \sum_{\tau=1}^{t} (x_{l,\tau} - x_{l+1,\tau}) \geq \sum_{\tau=1}^{t-1} (x_{l,\tau} - x_{l+1,\tau}) - x_{l+1,\tau} = (1 - p_t) \cdot \sum_{\tau=1}^{t-1} (x_{l,\tau} - x_{l+1,\tau}) \geq 0$$

Thus, we have proved that $0 \leq x_{l+1,t+1}$ and $\sum_{v=1}^{l+1} x_{v,t+1} \leq \theta \cdot p_{t+1}$. From the induction on t, we can conclude that for any $1 \leq t \leq T$, it holds that $0 \leq x_{l+1,t}$ and $\sum_{v=1}^{l+1} x_{v,t} \leq \theta \cdot p_t$. Again, from the

induction on l, we can conclude that for any $1 \le l \le k$ and any $1 \le t \le T$, it holds that $0 \le x_{l,t}$ and $\sum_{v=1}^{l} x_{v,t} \le \theta \cdot p_t$, which completes our proof. □

Now we are ready to prove Lemma A.0.1.

Proof of Lemma A.0.1:. When $\{x_{l,t}(\theta)\}$ is feasible to Dual($\boldsymbol{p}, k$) in (2.5), we get from constraint (2.5b) and (2.5c) that

$$x_{1,T}(\theta) \le p_T \cdot (1 - \sum_{t=1}^{T-1} x_{1,t}(\theta)) \text{ and } x_{l,T}(\theta) \le p_T \cdot \sum_{t=1}^{T-1} (x_{l-1,t}(\theta) - x_{l,t}(\theta)) \ \forall l = 2, \dots, k$$

Summing up the above inequalities, we get

$$\sum_{l=1}^{k} x_{l,T}(\theta) \le p_T \cdot (1 - \sum_{t=1}^{T-1} x_{k,t}(\theta))$$

Further note that by definition, we have $\sum_{l=1}^{k} x_{l,T}(\theta) = \theta \cdot p_T$. Thus, we show that $\{x_{l,t}(\theta)\}$ is feasible implies that $\sum_{t=1}^{T-1} x_{k,t}(\theta) \le 1 - \theta$.

Now we prove the reverse direction. Note that from the definition of $\{x_{l,t}(\theta)\}$, we have that $x_{l,t}(\theta) \le p_t \cdot \sum_{\tau=1}^{t-1} (x_{l-1,\tau}(\theta) - x_{l,\tau}(\theta))$ holds for any $1 \le l \le k-1$ and any $1 \le t \le T$, where we set $\sum_{\tau=1}^{t-1} x_{0,\tau}(\theta) = 1$ for any t for simplicity. Also, $\{x_{l,t}(\theta)\}$ are nonnegative as shown by Lemma A.0.3. Thus, we have that

$$\{x_{l,t}(\theta)\} \text{ is feasible } \Leftrightarrow x_{k,t}(\theta) \le p_t \cdot \sum_{\tau=1}^{t-1} (x_{k-1,\tau}(\theta) - x_{k,\tau}(\theta)) \text{ holds for any } t_k + 1 \le t \le T$$

Moreover, note that from definition, for $t_k + 1 \le t \le T$, we have that $x_{l,t}(\theta) = p_t \cdot \sum_{\tau=1}^{t-1} (x_{l-1,\tau}(\theta) - x_{l,\tau}(\theta))$ when $1 \le l \le k-1$. Thus, for $t_k + 1 \le t \le T$, we have that

$$x_{k,t}(\theta) \le p_t \cdot \sum_{\tau=1}^{t-1} (x_{k-1,\tau}(\theta) - x_{k,\tau}(\theta)) \Leftrightarrow \theta = \sum_{v=1}^{k} x_{v,t}/p_t \le 1 - \sum_{\tau=1}^{t-1} x_{k,\tau}(\theta)$$

From the nonnegativity of $\{x_{l,t}(\theta)\}$, we know that $\sum_{\tau=1}^{t-1} x_{k,\tau}(\theta)$ is monotone increasing with t. Thus, it holds that

$$\{x_{l,t}(\theta)\} \text{ is feasible } \Leftrightarrow \theta \leq 1 - \sum_{t=1}^{T-1} x_{k,t}(\theta)$$

which completes our proof. □

Proof of Lemma A.0.2:. For any fixed $\theta \in [0,1]$ and any fixed $\Delta \geq 0$ such that $\theta + \Delta \in [0,1]$, we compare between $\{x_{l,t}(\theta)\}$ and $\{x_{l,t}(\theta+\Delta)\}$. Since we consider for a fixed θ and Δ, for notation brevity, we will omit θ and Δ by substituting $\{x_{l,t}\}$ for $\{x_{l,t}(\theta)\}$ and substituting $\{x'_{l,t}\}$ for $\{x_{l,t}(\theta+\Delta)\}$. Respectively, we denote $y_{l,t} = \sum_{\tau=1}^{t-1} x_{l,\tau}$ and $y'_{l,t} = \sum_{\tau=1}^{t-1} x'_{l,\tau}$. Also, we denote $\{t_l\}$ to be the time indexes associated with $\{x_{l,t}\}$ in the definition of $\{x_{l,t}\}$ and $\{t'_l\}$ to be the time indexes associated with $\{x'_{l,t}\}$. We will use induction to show that for each l, we have that $y_{l,t} \leq y'_{l,t}$ and $\sum_{v=1}^{l} y'_{v,t} \leq \sum_{v=1}^{l} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ hold for each t.

For the case $l = 1$, obviously we have that $t'_2 \leq t_2$. When $1 \leq t \leq t'_2$, from definition, it holds that $y_{1,t} \leq y'_{1,t} \leq y_{1,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$. We now use induction on t for $t'_2 + 1 \leq t \leq t_2$. When $t = t'_2 + 1$, note that

$$x'_{1,t'_2+1} = p_{t'_2+1} \cdot (1 - y'_{1,t'_2}) \leq (\theta + \Delta) \cdot p_{t'_2+1} \text{ and } x_{1,t'_2+1} = \theta \cdot p_{t'_2+1} \leq p_{t'_2+1} \cdot (1 - y_{1,t'_2})$$

we have that

$$y_{1,t'_2+1} = y_{1,t'_2} + x_{1,t'_2+1} \leq p_{t'_2+1} + (1 - p_{t'_2+1}) \cdot y_{1,t'_2} \leq p_{t'_2+1} + (1 - p_{t'_2+1}) \cdot y'_{1,t'_2} = y'_{1,t'_2+1}$$

and

$$y'_{1,t'_2+1} = y'_{1,t'_2} + x'_{1,t'_2+1} \leq y_{1,t'_2} + \Delta \cdot \sum_{t=1}^{t'_2} p_t + (\theta + \Delta) \cdot p_{t'_2+1} = y_{1,t'_2+1} + \Delta \cdot \sum_{t=1}^{t'_2+1} p_t$$

Now suppose for a fixed t satisfying $t'_2 + 1 \leq t \leq t_2 - 1$, it holds $y_{1,t} \leq y'_{1,t} \leq y_{1,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$.

From definition, note that

$$x'_{1,t+1} = p_{t+1} \cdot (1 - y'_{1,t}) \leq (\theta + \Delta) \cdot p_{t+1} \text{ and } x_{1,t+1} = \theta \cdot p_{t+1} \leq p_{t+1} \cdot (1 - y_{1,t})$$

we have

$$y_{1,t+1} = y_{1,t} + x_{1,t+1} \leq p_{t+1} + (1 - p_{t+1}) \cdot y_{1,t} \leq p_{t+1} + (1 - p_{t+1}) \cdot y'_{1,t} = y'_{1,t+1}$$

and

$$y'_{1,t+1} = y'_{1,t} + x'_{1,t+1} \leq y_{1,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau + (\theta + \Delta) \cdot p_{t+1} = y_{1,t+1} + \Delta \cdot \sum_{\tau=1}^{t+1} p_\tau$$

Thus, from induction on t, we conclude that $y_{1,t} \leq y'_{1,t} \leq y_{1,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ holds for any $t'_2 + 1 \leq t \leq t_2$. Finally, when $t \geq t_2 + 1$, note that

$$y_{1,t} = p_t + (1 - p_t) \cdot y_{1,t-1} \text{ and } y'_{1,t} = p_t + (1 - p_t) \cdot y'_{1,t-1}$$

which implies that

$$y'_{1,t} - y_{1,t} = (1 - p_t) \cdot (y'_{1,t-1} - y_{1,t-1}) = \cdots = (y'_{1,t_2} - y_{1,t_2}) \cdot \prod_{\tau=t_2+1}^{t} (1 - p_\tau)$$

Thus, we prove that for any $1 \leq t \leq T$, it holds that $y_{1,t} \leq y'_{1,t} \leq y_{1,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$.

Suppose that for a fixed $1 \leq l \leq k$, $y_{l,t} \leq y'_{l,t}$ and $\sum_{v=1}^{l} y'_{v,t} \leq \sum_{v=1}^{l} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ hold for each t. We now consider the case for $l+1$. When $t \leq \min\{t_{l+2}, t'_{l+2}\}$, from definition, we have that

$$\sum_{v=1}^{l+1} y'_{v,t} = (\theta + \Delta) \cdot \sum_{\tau=1}^{t} p_\tau \text{ and } \sum_{v=1}^{l+1} y_{v,t} = \theta \cdot \sum_{\tau=1}^{t} p_\tau$$

which implies that $\sum_{v=1}^{l+1} y'_{v,t} \le \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$. Also, we have

$$y'_{l+1,t} - y_{l+1,t} = \Delta \cdot \sum_{\tau=1}^{t} p_\tau - \left(\sum_{v=1}^{l} y'_{v,t} - \sum_{v=1}^{l} y_{v,t} \right) \ge 0$$

where the last inequality holds from induction condition. Thus, we prove that $y_{l+1,t} \le y'_{l+1,t}$ and $\sum_{v=1}^{l+1} y'_{v,t} \le \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ hold for each $1 \le t \le \min\{t_{l+2}, t'_{l+2}\}$. Moreover, note that t_{l+2} is defined as the first time that $\theta > 1 - y_{l+1,t_{l+2}}$ while t'_{l+2} is defined as the first time that $\theta + \Delta > 1 - y'_{l+1,t'_{l+2}}$. Since $y'_{l+1,t} \ge y_{l+1,t}$ when $t \le \min\{t_{l+2}, t'_{l+2}\}$, we must have $t'_{l+2} \le t_{l+2}$. Then we use induction on t for $t'_{l+2} + 1 \le t \le t_{l+2}$. When $t = t'_{l+2} + 1 \le t_{l+2}$, from definition, we have

$$x'_{l+1,t'_{l+2}+1} = p_{t'_{l+2}+1} \cdot (y'_{l,t'_{l+2}} - y'_{l+1,t'_{l+2}}) \Rightarrow y'_{l+1,t'_{l+2}+1} = p_{t'_{l+2}+1} \cdot y'_{l,t'_{l+2}} + (1 - p_{t'_{l+2}+1}) \cdot y'_{l+1,t'_{l+2}}$$

and

$$x_{l+1,t'_{l+2}+1} \le p_{t'_{l+2}+1} \cdot (y_{l,t'_{l+2}} - y_{l+1,t'_{l+2}}) \Rightarrow y_{l+1,t'_{l+2}+1} \le p_{t'_{l+2}+1} \cdot y_{l,t'_{l+2}} + (1 - p_{t'_{l+2}+1}) \cdot y_{l+1,t'_{l+2}}$$

Note that $y'_{l,t'_{l+2}} \ge y_{l,t'_{l+2}}$ and $y'_{l+1,t'_{l+2}} \ge y_{l+1,t'_{l+2}}$, we get $y'_{l+1,t'_{l+2}+1} \ge y_{l+1,t'_{l+2}+1}$. Moreover, note that from the definition of t'_{l+2}, we have

$$\sum_{v=1}^{l+1} x'_{v,t'_{l+2}+1} \le p_{t'_{l+2}+1} \cdot (\theta + \Delta) = \sum_{v=1}^{l+1} x_{v,t'_{l+2}+1} + \Delta \cdot p_{t'_{l+2}+1}$$

which implies that

$$\begin{aligned} \sum_{v=1}^{l+1} y'_{v,t'_{l+2}+1} &= \sum_{v=1}^{l+1} y'_{v,t'_{l+2}} + \sum_{v=1}^{l+1} x'_{v,t'_{l+2}+1} \le \sum_{v=1}^{l+1} y_{v,t'_{l+2}} + \Delta \cdot \sum_{j=1}^{t'_{l+2}} p_j + \sum_{v=1}^{l+1} x_{v,t'_{l+2}+1} + \Delta \cdot p_{t'_{l+2}+1} \\ &= \sum_{v=1}^{l+1} y_{v,t'_{l+2}+1} + \Delta \cdot \sum_{j=1}^{t'_{l+2}+1} p_j \end{aligned}$$

Then suppose for a fixed t satisfying $t'_{l+2} + 1 \leq t \leq t_{l+2} - 1$, it holds that $y_{l+1,t} \leq y'_{l+1,t}$ and $\sum_{v=1}^{l+1} y'_{v,t} \leq \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$. From definition, we have

$$x'_{l+1,t+1} = p_{t+1} \cdot (y'_{l,t} - y'_{l+1,t}) \Rightarrow y'_{l+1,t+1} = p_{t+1} \cdot y'_{l,t} + (1 - p_{t+1}) \cdot y'_{l+1,t}$$

and

$$x_{l+1,t+1} \leq p_{t+1} \cdot (y_{l,t} - y_{l+1,t}) \Rightarrow y_{l+1,t+1} \leq p_{t+1} \cdot y_{l,t} + (1 - p_{t+1}) \cdot y_{l+1,t}$$

Note that $y'_{l,t} \geq y_{l,t}$ and $y'_{l+1,t} \geq y_{l+1,t}$, we have $y'_{l+1,t+1} \geq y_{l+1,t+1}$. Also, from the definition of t'_{l+2}, we have

$$\sum_{v=1}^{l+1} x'_{v,t+1} \leq p_{t+1} \cdot (\theta + \Delta) = \sum_{v=1}^{l+1} x_{v,t+1} + \Delta \cdot p_{t+1}$$

which implies that

$$\begin{aligned}\sum_{v=1}^{l+1} y'_{v,t+1} &= \sum_{v=1}^{l+1} y'_{v,t} + \sum_{v=1}^{l+1} x'_{v,t+1} \leq \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau + \sum_{v=1}^{l+1} x_{v,t+1} + \Delta \cdot p_{t+1} \\ &= \sum_{v=1}^{l+1} y_{v,t+1} + \Delta \cdot \sum_{\tau=1}^{t+1} p_\tau\end{aligned}$$

Thus, from induction on t, we prove that $y_{l+1,t} \leq y'_{l+1,t}$ and $\sum_{v=1}^{l+1} y'_{v,t} \leq \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ hold for any $t'_{l+2} + 1 \leq t \leq t_{l+2}$. Finally, when $t \geq t_{l+2} + 1$, note that

$$y_{l+1,t} = p_t \cdot y_{l,t-1} + (1 - p_t) \cdot y_{l+1,t-1} \text{ and } y'_{l+1,t} = p_t \cdot y'_{l,t-1} + (1 - p_t) \cdot y'_{l+1,t-1}$$

which implies that

$$y'_{l+1,t} - y_{l+1,t} = p_t \cdot (y'_{l,t-1} - y_{l,t-1}) + (1 - p_t) \cdot (y'_{l+1,t-1} - y_{l+1,t-1})$$

It is direct to show inductively on t such that $y_{l+1,t} \leq y'_{l+1,t}$ and $\sum_{v=1}^{l+1} y'_{v,t} \leq \sum_{v=1}^{l+1} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$

hold for any $t \geq t_{l+2} + 1$.

Thus, we have proved that for any $1 \leq t \leq T$, we have $y_{l+1,t} \leq y'_{l+1,t}$ and $\sum_{v=1}^{l+1} y'_{v,t} \leq \sum_{v=1}^{l} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$. By the induction on l, we finally prove that for any $1 \leq l \leq k$, $y_{l,t} \leq y'_{l,t}$ and $\sum_{v=1}^{l} y'_{v,t} \leq \sum_{v=1}^{l} y_{v,t} + \Delta \cdot \sum_{\tau=1}^{t} p_\tau$ hold for any $1 \leq t \leq T$. In this way, we prove that $y_{l,t}(\theta)$ is monotone increasing with θ for any l, t. Moreover, note that since $\sum_{\tau=1}^{T} p_\tau \leq k$, we have that $y_{l,t}(\theta + \Delta) \leq y_{l,t}(\theta) + k \cdot \Delta$ hold for any θ, Δ and any l, t. Thus, $y_{l,t}(\theta)$ is a continuous function on θ, which completes our proof. □

Construction of $\{\beta^*_{l,t}, \xi^*_t\}$ and Proof of Theorem 2.4

Given Lemma 2.3.1, in order to prove Theorem 2.4, it is enough for us to construct a feasible solution $\{\beta^*_{l,t}, \xi^*_t\}$ to Primal$(\boldsymbol{p}, k)$ such that the primal-dual pair $\{\theta^*, x_{l,t}(\theta^*)\}$ and $\{\beta^*_{l,t}, \xi^*_t\}$ satisfies the complementary slackness conditions. Specifically, we will construct a feasible solution $\{\beta^*_{l,t}, \xi^*_t\}$ to Primal$(\boldsymbol{p}, k)$ satisfying the following conditions:

$$\begin{aligned}
&\beta^*_{1,t} \cdot \left(x_{1,t}(\theta^*) - p_t \cdot (1 - \sum_{\tau<t} x_{1,\tau}(\theta^*)) \right) = 0, \quad \forall t = 1, \ldots, T \\
&\beta^*_{l,t} \cdot \left(x_{l,t}(\theta^*) - p_t \cdot \sum_{\tau<t} (x_{l-1,\tau}(\theta^*) - x_{l,\tau}(\theta^*)) \right) = 0, \quad \forall t = 1, \ldots, T, \forall l = 2, \ldots, k \\
&x_{l,t}(\theta^*) \cdot \left(\beta^*_{l,t} + \sum_{\tau>t} p_\tau \cdot (\beta^*_{l,\tau} - \beta^*_{l+1,\tau}) - \xi^*_t \right) = 0, \quad \forall t = 1, \ldots, T, \forall l = 2, \ldots, k \\
&x_{k,t}(\theta^*) \cdot \left(\beta^*_{k,t} + \sum_{\tau>t} p_\tau \cdot \beta^*_{k,\tau} - \xi^*_t \right) = 0, \quad \forall t = 1, \ldots, T
\end{aligned} \tag{A.4}$$

Note that from definitions, $\{x_{l,t}(\theta^*)\}$ satisfies the following conditions:

$$\begin{aligned}
&x_{l,t}(\theta^*) = 0 \leq p_t \cdot \sum_{\tau=1}^{t-1} (x_{l-1,\tau}(\theta^*) - x_{l,\tau}(\theta^*)), \quad \forall t \leq t_l \\
&x_{l,t}(\theta^*) = \theta^* \cdot p_t - \sum_{v=1}^{l-1} x_{v,t}(\theta^*) \leq p_t \cdot \sum_{\tau=1}^{t-1} (x_{l-1,\tau}(\theta^*) - x_{l,\tau}(\theta^*)) \quad \text{for } t_l + 1 \leq t \leq t_{l+1}
\end{aligned}$$

where $\{t_l\}$ are the time indexes associated with the definition of $\{x_{l,t}(\theta^*)\}$ and we define $t_1 = 0$, $t_{k+1} = T - 1$. For simplicity, we also denote $\sum_{\tau=1}^{t-1} x_{0,\tau}(\theta^*) = 1$ for any t. Thus, in order for $\{\beta^*_{l,t}, \xi^*_t\}$ to satisfy the conditions in (A.4), it is enough for $\{\beta^*_{l,t}, \xi^*_t\}$ to be feasible to Primal$(\boldsymbol{p}, k)$ and satisfy the following conditions:

$$\beta^*_{l,t} = 0 \quad \text{for } t \le t_{l+1} \tag{A.5}$$

$$\beta^*_{l,t} + \sum_{\tau=t+1}^{T} p_\tau \cdot (\beta^*_{l,\tau} - \beta^*_{l+1,\tau}) = \xi^*_t \quad \text{for } t \ge t_l + 1 \tag{A.6}$$

where we denote $\beta^*_{k+1,t} = 0$ for notation simplicity. We now show the construction of the solution $\{\beta^*_{l,t}, \xi^*_t\}$ to Primal$(\boldsymbol{p}, k)$. Define the following constants for each $l, q \in \{1, 2, \ldots, k\}$:

$$B_{l,q} = \sum_{t_l+1 \le j_1 < j_2 < \cdots < j_q \le t_{l+1}} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{w=t_l+1}^{t_{l+1}} (1-p_w)$$

and we set $B_{l,0} = \prod_{w=t_l+1}^{t_{l+1}} (1-p_w)$. We also define the following terms for each $l, q \in \{1, 2, \ldots, k\}$ and each $t \in \{t_l + 1, \ldots, t_{l+1}\}$, where $\{t_l\}$ are the time indexes defined in the construction of $\{\theta^*, x_{l,t}(\theta^*)\}$ and we define $t_1 = 0$, $t_{k+1} = T - 1$:

$$A_{l,q}(t) = \sum_{t+1 \le j_1 < j_2 < \cdots < j_q \le t_{l+1}} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{w=t+1}^{t_{l+1}} (1-p_w)$$

and we set $A_{l,0}(t) = \prod_{w=t+1}^{t_{l+1}}(1-p_w)$. Then our construction of the solution $\{\beta^*_{l,t}, \xi^*_t\}$ can be fully described as follows:

$$\begin{aligned}
&\xi^*_T = \beta^*_{l_1,T} = R, \quad \forall l_1 = 1,2,\ldots,k \\
&\beta^*_{l_1,t} = 0, \quad \forall l_1 = 1,2,\ldots,k, \forall t \le t_{l_1+1} \\
&\xi^*_t = \phi_l \cdot p_T R, \quad \forall l = 1,2,\ldots,k, \forall t_l + 1 \le t \le t_{l+1} \\
&\beta^*_{l_1,t} = p_T R \cdot \sum_{w=l_1}^{l_2-1} \delta_{w,l_2} \cdot A_{l_2,w-l_1}(t), \quad \forall l_1 = 1,2,\ldots,k, \forall l_2 = l_1+1,\ldots,k, \forall t_{l_2}+1 \le t \le t_{l_2+1}
\end{aligned} \tag{A.7}$$

where the parameters $\{\phi_l, \delta_{l_1,l_2}, R\}$ are defined as:

$$\delta_{l,k} = 1 \qquad \forall l = 1,2,\ldots,k-1$$

$$\delta_{l,l} = 0 \qquad \forall l = 1,2,\ldots,k$$

$$\delta_{l_1,l_2} = \sum_{w_0=l_1+1}^{l_2} \sum_{w_1=w_0}^{l_2+1} \sum_{w_2=w_1}^{l_2+2} \cdots \sum_{w_{k-1-l_2}=w_{k-2-l_2}}^{k-1} B_{l_2+1,w_1-w_0} \cdot B_{l_2+2,w_2-w_1} \ldots B_{k-1,w_{k-1-l_2}-w_{k-2-l_2}} \cdot B_{k,k-w_{k-1-l_2}},$$
$$\forall l_2 = 1,2,\ldots,k-1 \text{ and } l_1 = 1,2,\ldots,l_2-1$$

$$\phi_k = 1$$

$$\phi_l = \sum_{q=l+1}^{k} \sum_{w=l+1}^{q} (\delta_{w-1,q} - \delta_{w,q}) \cdot (1 - \sum_{v=0}^{w-l-1} B_{q,v}) \quad \forall l = 1,2,\ldots,k-1$$

and R is a positive constant such that $\sum_{t=1}^{T} p_t \cdot \xi^*_t = 1$. We then prove the feasibility of $\{\beta^*_{l,t}, \xi^*_t\}$ and the conditions (A.5), (A.6) are satisfied. Obviously, from definition, $\beta^*_{l,t}$ is nonnegative for each l and each t. We first prove that ξ^*_t is also nonnegative for each t.

Lemma A.0.4. *For each $l_2 = 1,2,\ldots,k$ and each $l_1 = 1,2,\ldots,l_2-1$, we have that $\delta_{l_1,l_2} \ge \delta_{l_1+1,l_2}$.*

Proof:. Note that when $l_2 = k$, we have that $\delta_{l,k} = 1$ for each $l = 1, 2, \ldots, k-1$, thus it holds that $\delta_{l,k} \geq \delta_{l+1,k}$. When $l_2 \leq k-1$, from definitions, we have that for each $l_1 = 1, 2, \ldots, l_2 - 1$

$$\delta_{l_1,l_2} - \delta_{l_1+1,l_2} = \sum_{w_1=l_1+1}^{l_2+1} \sum_{w_2=w_1}^{l_2+2} \cdots \sum_{w_{k-1-l_2}=w_{k-2-l_2}}^{k-1} B_{l_2+1,w_1-l_1-1} \cdot B_{l_2+2,w_2-w_1} \ldots B_{k-1,w_{k-1-l_2}-w_{k-2-l_2}} \cdot B_{k,k-w_{k-1-l_2}}$$

which completes our proof. □ We then show that the term $1 - \sum_{w=0}^{q} B_{l,w}$ is nonnegative for each l and each q. Note that the following lemma essentially implies that $\sum_{t=t_l+1}^{t_{l+1}} p_t \cdot A_{l,q}(t) = 1 - \sum_{w=0}^{q} B_{l,w}$, by replacing i_1 with t_l and i_2 with t_{l+1} in (A.8), which establishes the nonnegativity of the term $1 - \sum_{w=0}^{q} B_{l,w}$.

Lemma A.0.5. *For each $q \in \{1, 2, \ldots, k\}$ and any $1 \leq i_1 + 1 \leq i_2 \leq T$, it holds that*

$$\begin{aligned} &\sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{t+1 \leq j_1 < j_2 < \cdots < j_q \leq i_2} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2} (1-p_v) \\ &= 1 - \sum_{w=0}^{q} \sum_{i_1+1 \leq j_1 < j_2 < \cdots < j_w \leq i_2} \frac{p_{j_1} p_{j_2} \cdots p_{j_w}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_w})} \cdot \prod_{v=i_1+1}^{i_2} (1-p_v), \end{aligned} \tag{A.8}$$

Proof:. We will do induction on q from $q = 0$ to $q = k$ to prove (A.8). When $q = 0$, we have that

$$\begin{aligned} \sum_{t=i_1+1}^{i_2} p_t \cdot \prod_{v=t+1}^{i_2} (1-p_v) &= \sum_{t=i_1+1}^{i_2} (1-(1-p_t)) \cdot \prod_{v=t+1}^{i_2} (1-p_v) = \sum_{t=i_1+1}^{i_2} \left(\prod_{v=t+1}^{i_2} (1-p_v) - \prod_{v=t}^{i_2} (1-p_v) \right) \\ &= 1 - \prod_{v=i_1+1}^{i_2} (1-p_v) \end{aligned}$$

Thus, we have (A.8) holds for $q = 0$. Suppose (A.8) holds for $1, 2, \ldots, q-1$, we consider the case for q. For any $1 \le i_1 + 1 \le i_2 \le T$, we have that

$$
\begin{aligned}
&\sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{t+1 \le j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2} (1-p_v) \\
&= \sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{j_1=t+1}^{i_2} \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2} (1-p_v) \\
&= \sum_{j_1=i_1+2}^{i_2} \sum_{t=i_1+1}^{j_1-1} p_t \cdot \frac{p_{j_1}}{1-p_{j_1}} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2} (1-p_v) \\
&= \sum_{j_1=i_1+2}^{i_2} \frac{p_{j_1}}{1-p_{j_1}} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=j_1}^{i_2} (1-p_v) \cdot \sum_{t=i_1+1}^{j_1-1} p_t \cdot \prod_{v=t+1}^{j_1-1} (1-p_v)
\end{aligned}
$$

where the second equality holds by exchanging the order of summation. Note that for induction purpose, we assume (A.8) holds for $q = 0$, which implies that $\sum_{t=i_1+1}^{j_1-1} p_t \cdot \prod_{v=t+1}^{j_1-1}(1-p_v) = 1 - \prod_{v=i_1+1}^{j_1-1}(1-p_v)$. Then we have

$$
\begin{aligned}
&\sum_{j_1=i_1+2}^{i_2} \frac{p_{j_1}}{1-p_{j_1}} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=j_1}^{i_2} (1-p_v) \cdot \sum_{t=i_1+1}^{j_1-1} p_t \cdot \prod_{v=t+1}^{j_1-1} (1-p_v) \\
&= \sum_{j_1=i_1+2}^{i_2} p_{j_1} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2} (1-p_v) \cdot \left(1 - \prod_{v=i_1+1}^{j_1-1} (1-p_v)\right) \\
&= \sum_{j_1=i_1+1}^{i_2} p_{j_1} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2} (1-p_v) \cdot \left(1 - \prod_{v=i_1+1}^{j_1-1} (1-p_v)\right)
\end{aligned}
$$

where the second equality holds by noting that when $j_1 = i_1 + 1$, we have $1 - \prod_{v=i_1+1}^{j_1-1}(1-p_v) = 0$. Thus, it holds that

$$
\begin{aligned}
&\sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{t+1 \le j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_1} p_{j_2} \cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2} (1-p_v) \\
&= \sum_{j_1=i_1+1}^{i_2} p_{j_1} \cdot \sum_{j_1 < j_2 < \cdots < j_q \le i_2} \frac{p_{j_2} \cdots p_{j_q}}{(1-p_{j_2}) \ldots (1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2} (1-p_v) \cdot \left(1 - \prod_{v=i_1+1}^{j_1-1} (1-p_v)\right)
\end{aligned}
$$

Note that for the induction purpose, we assume that (A.8) holds for $q-1$. Then, we have that

$$\begin{aligned}
&\sum_{j_1=i_1+1}^{i_2} p_{j_1} \cdot \sum_{j_1<j_2<\cdots<j_q\le i_2} \frac{p_{j_2}\cdots p_{j_q}}{(1-p_{j_2})\ldots(1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2}(1-p_v) \\
&= \sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{t+1\le j_1<j_2<\cdots<j_{q-1}\le i_2} \frac{p_{j_1}p_{j_2}\cdots p_{j_{q-1}}}{(1-p_{j_1})(1-p_{j_2})\ldots(1-p_{j_{q-1}})} \cdot \prod_{v=t+1}^{i_2}(1-p_v) \\
&= 1-\sum_{w=0}^{q-1} \sum_{i_1+1\le j_1<j_2<\cdots<j_w\le i_2} \frac{p_{j_1}p_{j_2}\cdots p_{j_w}}{(1-p_{j_1})(1-p_{j_2})\ldots(1-p_{j_w})} \cdot \prod_{v=i_1+1}^{i_2}(1-p_v)
\end{aligned}$$

where the second equality holds from replacing the index j_{l+1} with j_l for $l=2,\ldots,q$ and replace the index j_1 with t. Also, note that

$$\begin{aligned}
&\sum_{j_1=i_1+1}^{i_2} p_{j_1} \cdot \sum_{j_1<j_2<\cdots<j_q\le i_2} \frac{p_{j_2}\cdots p_{j_q}}{(1-p_{j_2})\ldots(1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2}(1-p_v) \cdot \prod_{v=i_1+1}^{j_1-1}(1-p_v) \\
&= \sum_{i_1+1\le j_1<j_2<\cdots<j_q\le i_2} \frac{p_{j_1}p_{j_2}\cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2})\ldots(1-p_{j_q})} \cdot \prod_{v=i_1+1}^{i_2}(1-p_v)
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
&\sum_{t=i_1+1}^{i_2} p_t \cdot \sum_{t+1\le j_1<j_2<\cdots<j_q\le i_2} \frac{p_{j_1}p_{j_2}\cdots p_{j_q}}{(1-p_{j_1})(1-p_{j_2})\ldots(1-p_{j_q})} \cdot \prod_{v=t+1}^{i_2}(1-p_v) \\
&= \sum_{j_1=i_1+1}^{i_2} p_{j_1} \cdot \sum_{j_1<j_2<\cdots<j_q\le i_2} \frac{p_{j_2}\cdots p_{j_q}}{(1-p_{j_2})\ldots(1-p_{j_q})} \cdot \prod_{v=j_1+1}^{i_2}(1-p_v) \cdot \left(1-\prod_{v=i_1+1}^{j_1-1}(1-p_v)\right) \\
&= 1-\sum_{w=0}^{q} \sum_{i_1+1\le j_1<j_2<\cdots<j_w\le i_2} \frac{p_{j_1}p_{j_2}\cdots p_{j_w}}{(1-p_{j_1})(1-p_{j_2})\ldots(1-p_{j_w})} \cdot \prod_{v=i_1+1}^{i_2}(1-p_v)
\end{aligned}$$

which completes our proof by induction on q. □

Combining Lemma A.0.4 and Lemma A.0.5, we draw the following conclusion.

Lemma A.0.6. *For each $l=1,2,\ldots,k$ and each $t=1,2,\ldots,T$, we have that $\beta^*_{l,t}\ge 0$ and $\xi^*_t\ge 0$.*

Proof:. Note that from definition, $\beta^*_{l,t} \geq 0$ for each l and t. We then show the non-negativity of ξ^*_t for each t. Note that Lemma A.0.4 shows that $\delta_{l_1,l_2} \geq \delta_{l_1+1,l_2}$ for each $l_2 = 1, 2, \ldots, k$ and each $l_1 = 1, 2, \ldots, l_2 - 1$. It only remains to show the non-negativity of the term $1 - \sum_{w=0}^{q} B_{l,w}$, which can be directly established by Lemma A.0.5. Specifically, by replacing i_1 with t_l and i_2 with t_{l+1} in (A.8), we have $1 - \sum_{w=0}^{q} B_{l,w} = \sum_{t=t_l+1}^{t_{l+1}} p_t \cdot A_{l,q}(t) \geq 0$. □

From the definition of $\{\beta^*_{l,t}, \xi^*_t\}$, condition (A.5) holds obviously. We then prove that condition (A.6) is satisfied.

Lemma A.0.7. *For each $l = 1, 2, \ldots, k$ and each $t \geq t_l + 1$, it holds that*

$$\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = \xi^*_t$$

where we denote $\beta^*_{k+1,t} = 0$ for notation simplicity.

Proof:. When $l = k$, from definition, we have $\beta^*_{l,j} = 0$ for each $j \leq t_{l+1} = T - 1$ and $\beta^*_{l,T} = R$, thus the lemma holds directly. When $t = T$, it is also direct to show from definition that the lemma holds. We then focus on the case where $l \leq k - 1$ and $t \leq T - 1$.

For a fixed $l \leq k - 1$ and a fixed $t_l + 1 \leq t \leq T - 1$, we denote an index $l_1 \geq l$ such that $t_{l_1} + 1 \leq t \leq t_{l_1+1}$. We then consider the following cases separately based on the value of l_1.

(i). When $l_1 \leq k - 1$, we have that

$$\beta^*_{l,t} = p_T R \cdot \sum_{w=l}^{l_1-1} \delta_{w,l_1} \cdot A_{l_1,w-l}(t) \tag{A.9}$$

also, for any $t + 1 \leq j \leq t_{l_1+1}$, we have that

$$\beta^*_{l,j} - \beta^*_{l+1,j} = p_T R \cdot \sum_{w=l}^{l_1-1} (\delta_{w,l_1} - \delta_{w+1,l_1}) \cdot A_{l_1,w-l}(j)$$

which implies that

$$\sum_{j=t+1}^{t_{l_1+1}} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = p_T R \cdot \sum_{w=l}^{l_1-1} (\delta_{w,l_1} - \delta_{w+1,l_1}) \cdot \sum_{j=t+1}^{t_{l_1+1}} p_j \cdot A_{l_1,w-l}(j)$$

Note that from (A.8), it holds that $\sum_{j=t+1}^{t_{l_1+1}} p_j \cdot A_{l_1,w-l}(j) = 1 - \sum_{q=0}^{w-l} A_{l_1,q}(t)$. Thus, we have that

$$\begin{aligned}\sum_{j=t+1}^{t_{l_1+1}} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) &= p_T R \cdot \sum_{w=l}^{l_1-1} (\delta_{w,l_1} - \delta_{w+1,l_1}) \cdot \left(1 - \sum_{q=0}^{w-l} A_{l_1,q}(t)\right) \\ &= p_T R \cdot \delta_{l,l_1} - p_T R \cdot \sum_{w=l}^{l_1-1} \delta_{w,l_1} \cdot A_{l_1,w-l}(t)\end{aligned} \tag{A.10}$$

where the last equality holds from $\delta_{l_1,l_1} = 0$. Similarly, for any $l_2 \geq l_1 + 1$ and any $t_{l_2}+1 \leq j \leq t_{l_2+1}$, we have that

$$\beta^*_{l,j} - \beta^*_{l+1,j} = p_T R \cdot \sum_{w=l}^{l_2-1} (\delta_{w,l_2} - \delta_{w+1,l_2}) \cdot A_{l_2,w-l}(j)$$

which implies that

$$\sum_{j=t_{l_2}+1}^{t_{l_2+1}} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = p_T R \cdot \sum_{w=l}^{l_2-1} (\delta_{w,l_2} - \delta_{w+1,l_2}) \cdot \sum_{j=t_{l_2}+1}^{t_{l_2+1}} p_j \cdot A_{l_2,w-l}(j)$$

Note that from Lemma A.0.5, we have that $\sum_{j=t_{l_2}+1}^{t_{l_2+1}} p_j \cdot A_{l_2,w-l}(j) = 1 - \sum_{q=0}^{w-l} B_{l_2,q}$. Thus, we have that

$$\begin{aligned}\sum_{j=t_{l_1+1}+1}^{t_{k+1}} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) &= \sum_{l_2=l_1+1}^{k} \sum_{j=t_{l_2}+1}^{t_{l_2+1}} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) \\ &= \sum_{l_2=l_1+1}^{k} p_T R \cdot \sum_{w=l}^{l_2-1} (\delta_{w,l_2} - \delta_{w+1,l_2}) \cdot (1 - \sum_{q=0}^{w-l} B_{l_2,q}) \\ &= p_T R \cdot \sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l-1} B_{l_2,q})\end{aligned} \tag{A.11}$$

Combining (A.9), (A.10) and (A.11), we have that

$$\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = p_T R \cdot \delta_{l,l_1} + p_T R \cdot \sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l-1} B_{l_2,q}) \quad \text{(A.12)}$$

Note that

$$\xi^*_t = p_T R \cdot \sum_{l_2=l_1+1}^{k} \sum_{w=l_1+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l_1-1} B_{l_2,q})$$

in order to show $\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = \xi^*_t$, it is enough to prove that

$$\delta_{l,l_1} + \sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l-1} B_{l_2,q}) = \sum_{l_2=l_1+1}^{k} \sum_{w=l_1+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l_1-1} B_{l_2,q}) \quad \text{(A.13)}$$

Further note that

$$\begin{aligned}
&\sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l-1} B_{l_2,q}) \\
&= \sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) - \sum_{l_2=l_1+1}^{k} \sum_{w=l+1}^{l_2} \sum_{q=0}^{w-l-1} B_{l_2,q} \cdot (\delta_{w-1,l_2} - \delta_{w,l_2}) \\
&= \sum_{l_2=l_1+1}^{k} \delta_{l,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l-1} B_{l_2,q} \cdot \delta_{q+l,l_2}
\end{aligned}$$

and similarly, note that

$$\sum_{l_2=l_1+1}^{k} \sum_{w=l_1+1}^{l_2} (\delta_{w-1,l_2} - \delta_{w,l_2}) \cdot (1 - \sum_{q=0}^{w-l_1-1} B_{l_2,q}) = \sum_{l_2=l_1+1}^{k} \delta_{l_1,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l_1-1} B_{l_2,q} \cdot \delta_{q+l_1,l_2}$$

in order to prove (A.13), it is enough to show that

$$\sum_{l_2=l_1}^{k} \delta_{l,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l-1} B_{l_2,q} \cdot \delta_{q+l,l_2} = \sum_{l_2=l_1+1}^{k} \delta_{l_1,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l_1-1} B_{l_2,q} \cdot \delta_{q+l_1,l_2} \quad \text{(A.14)}$$

When $l_1 = l$, it is direct to check that (A.14) holds. The proof of (A.14) when $l_1 \geq l+1$ is relegated to Lemma A.0.8. Thus, we prove that when $l_1 \leq k-1$, it holds that $\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = \xi^*_t$.

(ii). When $l_1 = k$, we have that

$$\beta^*_{l,t} = p_T R \cdot \sum_{w=l}^{k-1} A_{k,w-l}(t)$$

and for each $t+1 \leq j \leq T-1$, it holds that

$$\beta^*_{l,j} - \beta^*_{l+1,j} = p_T R \cdot \left(\sum_{w=l}^{k-1} A_{k,w-l}(j) - \sum_{w=l+1}^{k-1} A_{k,w-l-1}(j) \right) = p_T R \cdot A_{k,k-1-l}(j)$$

Note that $\beta^*_{l,T} = \beta^*_{l+1,T} = R$, we have

$$\beta^*_{l,t} + \sum_{j=t+1}^{T-1} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = p_T R \cdot \left(\sum_{w=l}^{k-1} A_{k,w-l}(t) + \sum_{j=t+1}^{T-1} p_j \cdot A_{k,k-1-l}(j) \right)$$

Note that from (A.8), it holds that $\sum_{j=t+1}^{T-1} p_j \cdot A_{k,k-1-l}(j) = 1 - \sum_{q=0}^{k-1-l} A_{k,q}(t)$. Thus, we have that

$$\beta^*_{l,t} + \sum_{j=t+1}^{T-1} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) = p_T R = \xi^*_t$$

which completes our proof. □

Lemma A.0.8. *For each $l = 1, 2, \ldots, k-1$ and each $l_1 = l, l+1, \ldots, k-1$, it holds that*

$$\sum_{l_2=l_1}^{k} \delta_{l,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l-1} B_{l_2,q} \cdot \delta_{q+l,l_2} = \sum_{l_2=l_1+1}^{k} \delta_{l_1,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l_1-1} B_{l_2,q} \cdot \delta_{q+l_1,l_2} \tag{A.15}$$

Proof:. We now prove (A.15) by induction on l from $l = k-1$ to $l = 1$. When $l = k-1$, we must have $l_1 = k-1 = l$, then (A.15) holds obviously. Suppose that there exists a $1 \leq l' \leq k-2$ such that for any l satisfying $l'+1 \leq l \leq k-1$, (A.15) holds for each l_1 such that $l \leq l_1 \leq k-1$,

then we consider the case when $l = l'$. For this case, we again use induction on l_1 from $l_1 = k-1$ to $l_1 = l+1 = l'+1$. When $l_1 = k-1$, we have that

$$\sum_{l_2=l_1}^{k} \delta_{l,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l-1} B_{l_2,q} \cdot \delta_{q+l,l_2} = \delta_{l,k-1} + \delta_{l,k} - \sum_{q=0}^{k-l-1} B_{k,q} \cdot \delta_{q+l,k}$$

and

$$\sum_{l_2=l_1+1}^{k} \delta_{l_1,l_2} - \sum_{l_2=l_1+1}^{k} \sum_{q=0}^{l_2-l_1-1} B_{l_2,q} \cdot \delta_{q+l_1,l_2} = \delta_{k-1,k} - B_{k,0} \cdot \delta_{k-1,k}$$

Further note that from definition, $\delta_{v,k} = 1$ for each $v \leq k-1$ and $\delta_{l,k-1} = \sum_{w_0=l+1}^{k-1} B_{k,k-w_0} = \sum_{q=1}^{k-l-1} B_{k,q}$, it is obvious that (A.15) holds when $l_1 = k-1$. Now suppose that (A.15) holds for l_1+1 (we assume $l_1 \geq l+1$ since when $l_1 = l$, it is direct from definition that (A.15) holds), we consider the case for l_1. Note that

$$\text{LHS of (A.15)} = \delta_{l,l_1} - \sum_{q=0}^{l_1-l} B_{l_1+1,q} \cdot \delta_{q+l,l_1+1} + \sum_{l_2=l_1+1}^{k} \delta_{l,l_2} - \sum_{l_2=l_1+2}^{k} \sum_{q=0}^{l_2-l-1} B_{l_2,q} \cdot \delta_{q+l,l_2}$$

and

$$\text{RHS of (A.15)} = \delta_{l_1,l_1+1} - \sum_{l_2=l_1+1}^{k} B_{l_2,l_2-l_1-1} \cdot \delta_{l_2-1,l_2} + \sum_{l_2=l_1+2}^{k} \delta_{l_1,l_2} - \sum_{l_2=l_1+2}^{k} \sum_{q=0}^{l_2-l_1-2} B_{l_2,q} \cdot \delta_{q+l_1,l_2}$$

Since we suppose for induction that (A.15) holds for l_1+1, we have that

$$\text{(A.15) holds for } (l,l_1) \Leftrightarrow \delta_{l,l_1} - \sum_{q=0}^{l_1-l} B_{l_1+1,q} \cdot \delta_{q+l,l_1+1} = \delta_{l_1,l_1+1} - \sum_{l_2=l_1+1}^{k} B_{l_2,l_2-l_1-1} \cdot \delta_{l_2-1,l_2}$$

Further note that we have supposed for induction that (A.15) holds for $(l+1,l_1)$, which implies

$$\delta_{l+1,l_1} - \sum_{q=0}^{l_1-l-1} B_{l_1+1,q} \cdot \delta_{q+l+1,l_1+1} = \delta_{l_1,l_1+1} - \sum_{l_2=l_1+1}^{k} B_{l_2,l_2-l_1-1} \cdot \delta_{l_2-1,l_2}$$

Thus, it holds that

$$(\text{A.15}) \text{ holds for } (l, l_1) \Leftrightarrow \delta_{l,l_1} - \delta_{l+1,l_1} = \sum_{q=0}^{l_1-l} B_{l_1+1,q} \cdot (\delta_{q+l,l_1+1} - \delta_{q+l+1,l_1+1})$$

Finally, from definition, we have

$$\delta_{l,l_1} - \delta_{l+1,l_1} = \sum_{w_1=l+1}^{l_1+1} \sum_{w_2=w_1}^{l_1+2} \cdots \sum_{w_{k-1-l_1}=w_{k-2-l_1}}^{k-1} B_{l_1+1,w_1-l-1} \cdot B_{l_1+2,w_2-w_1} \ldots B_{k-1,w_{k-1-l_1}-w_{k-2-l_1}} \cdot B_{k,k-w_{k-1-l_1}}$$

and

$$\delta_{q+l,l_1+1} - \delta_{q+l+1,l_1+1} = \sum_{w_2=q+l+1}^{l_1+2} \cdots \sum_{w_{k-1-l_1}=w_{k-2-l_1}}^{k-1} B_{l_1+2,w_2-q-l-1} \ldots B_{k-1,w_{k-1-l_1}-w_{k-2-l_1}} \cdot B_{k,k-w_{k-1-l_1}}$$

which implies that

$$\delta_{l,l_1} - \delta_{l+1,l_1} = \sum_{q=0}^{l_1-l} B_{l_1+1,q} \cdot (\delta_{q+l,l_1+1} - \delta_{q+l+1,l_1+1}) \tag{A.16}$$

Thus, from induction, we prove that (A.15) holds for each $l_1 \geq l + 1$. Note that (A.15) holds obviously for $l_1 = l$, (A.15) holds for each $l_1 \geq l$. From the induction on l, we know that (A.15) holds for each $1 \leq l \leq k - 1$ and each $l \leq l_1 \leq k - 1$, which completes our proof. □

Finally, we only need to prove feasibility of $\{\beta^*_{l,t}, \xi^*_t\}$ in the following lemma.

Lemma A.0.9. *For each $l = 1, 2, \ldots, k$ and each $t = 1, 2, \ldots, t_l$, it holds that*

$$\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) \geq \xi^*_t$$

where we denote $\beta^*_{k+1,t} = 0$ for notation simplicity.

Proof:. Note that from Lemma A.0.4, we have $\delta_{w,l_2} \geq \delta_{w+1,l_2}$, which implies that $\beta^*_{l,j} \geq \beta^*_{l+1,j}$ for each l and j. Thus, we have that for each $t = 1, 2, \ldots, t_l$, it holds that

$$\beta^*_{l,t} + \sum_{j=t+1}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) \geq \beta^*_{l,t_l+1} + \sum_{j=t_l+2}^{T} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j})$$

Further note that Lemma A.0.7 implies that

$$\beta^*_{l,t_l+1} + \sum_{j=t_l+2}^{n} p_j \cdot (\beta^*_{l,j} - \beta^*_{l+1,j}) \geq \xi^*_{t_l+1}$$

Thus, it is enough to show that $\xi^*_i \leq \xi^*_{t_l+1}$ for each $t = 1, 2, \ldots, t_l$. From the definition of ξ^*_i, it is enough to show that $\phi_l \leq \phi_{l+1}$. When $l = k-1$, we have $\phi_{l+1} = \phi_k = 1$ and $\phi_l = \phi_{k-1} = 1 - B_{k,0}$, which implies that $\phi_{k-1} \leq \phi_k$. When $l \leq k-2$, from definition, we have

$$\phi_l - \phi_{l+1} = \sum_{q=l+1}^{k} (\delta_{l,q} - \delta_{l+1,q}) \cdot (1 - B_{q,0}) - \sum_{q=l+2}^{k} \sum_{w=l+2}^{q} (\delta_{w-1,q} - \delta_{w,q}) \cdot B_{q,w-l-1}$$

Note that in the proof of Lemma A.0.8, we proved (A.16), then when $k-1 \geq q \geq l+1$, we have

$$\delta_{l,q} - \delta_{l+1,q} = \sum_{w=0}^{q-l} B_{q+1,w} \cdot (\delta_{w+l,q+1} - \delta_{w+l+1,q+1}) = \sum_{w=l+1}^{q+1} B_{q+1,w-l-1} \cdot (\delta_{w-1,q+1} - \delta_{w,q+1})$$

Thus, it holds that

$$\begin{aligned}
\phi_l - \phi_{l+1} &= -\sum_{q=l+1}^{k-1} (\delta_{l,q} - \delta_{l+1,q}) \cdot B_{q,0} + \sum_{q=l+1}^{k-1} \sum_{w=l+1}^{q+1} B_{q+1,w-l-1} \cdot (\delta_{w-1,q+1} - \delta_{w,q+1}) \\
&\quad - \sum_{q=l+2}^{k} \sum_{w=l+2}^{q} (\delta_{w-1,q} - \delta_{w,q}) \cdot B_{q,w-l-1} \\
&= -\sum_{q=l+1}^{k-1} (\delta_{l,q} - \delta_{l+1,q}) \cdot B_{q,0} + \sum_{q=l+1}^{k-1} B_{q+1,0} \cdot (\delta_{l,q+1} - \delta_{l+1,q+1}) \\
&= -B_{l+1,0} \cdot \delta_{l,l+1} \leq 0
\end{aligned}$$

which completes our proof. □

Together, Lemma A.0.6, Lemma A.0.7, and Lemma A.0.9 establish the feasibility of $\{\beta^*_{l,t}, \xi^*_t\}$. Then, from the definition of $\{\beta^*_{l,t}, \xi^*_t\}$, obviously condition (A.5) is satisfied and from Lemma A.0.7, condition (A.6) is satisfied. Thus, we finish the proof of Theorem 2.4.

Proof of Lemma 2.4.1:. Since we have $\text{Dual}(\boldsymbol{p}, k) = \text{Primal}(\boldsymbol{p}, k)$, it is enough to consider the primal LP $\text{Primal}(\boldsymbol{p}, k)$ in (??) and prove that $\text{Primal}(\boldsymbol{p}, k) \geq \text{Primal}(\tilde{\boldsymbol{p}}, k)$. Suppose the optimal solution of $\text{Primal}(\boldsymbol{p}, k)$ is denoted as $\{\beta^*_{l,t}, \xi^*_t\}$, as constructed in (A.7), we then construct a feasible solution $\{\tilde{\beta}_{l,t}, \tilde{\xi}_t\}$ to $\text{Primal}(\tilde{\boldsymbol{p}}, k)$ as follows:

$$
\begin{aligned}
&\tilde{\xi}_t = \xi^*_t \quad \forall 1 \leq t < q, \quad \tilde{\xi}_q = \tilde{\xi}_{q+1} = \xi^*_q, \quad \tilde{\xi}_{t+1} = \xi^*_t \quad \forall q+1 \leq t \leq T \\
&\tilde{\beta}_{l,t} = \beta^*_{l,t} \quad \forall l = 1, \ldots, k, \forall 1 \leq t < q \\
&\tilde{\beta}_{l,q} = \tilde{\beta}_{l,q+1} = \beta^*_{l,q} \quad \forall l = 1, \ldots, k \\
&\tilde{\beta}_{l,t+1} = \beta^*_{l,t} \quad \forall l = 1, \ldots, k, \forall q+1 \leq t \leq T
\end{aligned}
$$

Note that we have

$$
\text{Primal}(\boldsymbol{p}, k) = \sum_{t=1}^{T} p_t \cdot \beta^*_{1,t} = \sum_{t=1}^{T+1} \tilde{p}_t \cdot \tilde{\beta}_{1,t}
$$

it is enough to prove that $\{\tilde{\beta}_{l,t}, \tilde{\xi}_t\}$ is feasible to $\text{Primal}(\tilde{\boldsymbol{p}}, k)$. Obviously, we have $\{\tilde{\beta}_{l,t}, \tilde{\xi}_t\}$ are non-negative and $\sum_{t=1}^{T+1} \tilde{p}_t \cdot \tilde{\xi}_t = \sum_{t=1}^{T} p_t \cdot \xi^*_t = 1$, then we only need to check whether the following constraint is satisfied:

$$
\tilde{\beta}_{l,t} + \sum_{\tau > t} \tilde{p}_\tau \cdot (\tilde{\beta}_{l,\tau} - \tilde{\beta}_{l+1,\tau}) - \tilde{\xi}_t \geq 0, \quad \forall l = 1, \ldots, k, \forall t = 1, \ldots, T+1 \tag{A.17}
$$

where we denote $\tilde{\beta}_{k+1,t} = 0$ for notation simplicity. Note that when $t \geq q+1$, we have that

$$
\tilde{\beta}_{l,t} + \sum_{\tau > t} \tilde{p}_\tau \cdot (\tilde{\beta}_{l,\tau} - \tilde{\beta}_{l+1,\tau}) - \tilde{\xi}_t = \beta^*_{l,t} + \sum_{\tau > t} p_\tau \cdot (\beta^*_{l,\tau} - \beta^*_{l+1,\tau}) - \xi^*_t \geq 0, \quad \forall l = 1, \ldots, k
$$

and when $1 \le t \le q-1$, we also have

$$\tilde{\beta}_{l,t} + \sum_{\tau>t} \tilde{p}_\tau \cdot (\tilde{\beta}_{l,\tau} - \tilde{\beta}_{l+1,\tau}) - \tilde{\xi}_t = \beta^*_{l,t} + \sum_{\tau>t} p_\tau \cdot (\beta^*_{l,\tau} - \beta^*_{l+1,\tau}) - \xi^*_t \ge 0, \quad \forall l = 1, \dots, k$$

by noting $\tilde{p}_q + \tilde{p}_{q+1} = p_q$. Now we consider the case when $t = q$, then for each $l = 1, \dots, k$, we have

$$\begin{aligned}
\tilde{\beta}_{l,q} + \sum_{j=q+1}^{T+1} \tilde{p}_j \cdot (\tilde{\beta}_{l-1,j} - \tilde{\beta}_{l,j}) - \tilde{\xi}_q &= \tilde{\beta}_{l,q} + \sum_{j=q+2}^{T+1} \tilde{p}_j \cdot (\tilde{\beta}_{l-1,j} - \tilde{\beta}_{l,j}) - \tilde{\xi}_q + \tilde{p}_{q+1} \cdot (\tilde{\beta}_{l-1,q+1} - \tilde{\beta}_{l,q+1}) \\
&= \beta^*_{l,q} + \sum_{j=q+1}^{T} p_j \cdot (\beta^*_{l-1,j} - \beta^*_{l,j}) - \xi^*_q + p_q \cdot (1-\sigma) \cdot (\beta^*_{l-1,q} - \beta^*_{l,q}) \\
&\ge p_q \cdot (1-\sigma) \cdot (\beta^*_{l-1,q} - \beta^*_{l,q})
\end{aligned}$$

Thus, it is enough to show that $\beta^*_{l-1,q} \ge \beta^*_{l,q}$ to prove feasibility. Note that from Lemma A.0.4, for each $l_2 = 1, 2, \dots, k$ and each $l_1 = 1, 2, \dots, l_2 - 1$, we have $\delta_{l_1,l_2} \ge \delta_{l_1+1,l_2}$, then, it is direct to show that $\beta^*_{l-1,q} \ge \beta^*_{l,q}$ from the construction (A.7), which completes our proof. □

Proof of Proposition 2.5:. We consider the following problem instance $\mathcal{H}$. At the beginning, there are two queries arriving deterministically with a reward 1. Then, over the time interval $[0, 1]$, there are queries with reward $r_1 > 1$ arriving according to a Poisson process with rate λ. At last, there is one query with a reward $\frac{r_2}{\epsilon}$ arriving with a probability ϵ for some small $\epsilon > 0$.

Obviously, since $r_1 > 1$ and ϵ is set to be small, the prophet will first serve the last query as long as it arrives, and then serve the queries with a reward r_1 as much as possible, and at least serve the first two queries. Then, we have that

$$\mathbb{E}_{\boldsymbol{I}\sim\boldsymbol{F}}[V^{\text{off}}(\boldsymbol{I})] = \hat{V} := r_2 + 2 \cdot \exp(-\lambda) + (r_1 + 1) \cdot \lambda \cdot \exp(-\lambda) + 2r_1 \cdot (1 - (\lambda + 1) \cdot \exp(-\lambda) + O(\epsilon)$$

Moreover, for any online algorithm π, we consider the following situations separately based on

the number of the first two queries that π will serve.

(i). If π will always serve the first two queries, then it is obvious that $\mathbb{E}_{\pi,\boldsymbol{I}\sim\boldsymbol{F}}[V^{\pi}(\boldsymbol{I})] = 2$.

(ii). If π serves only one of the first two queries, then the optimal way for π to serve the second query will depend on the value of r_1 and r_2. To be more specific, if $r_1 \geq r_2$, then the optimal way is to serve the query with reward r_1 as long as it arrives, and if $r_1 < r_2$, then the optimal way is to reject all the arriving queries with reward r_1 and only serve the last query. Thus, it holds that

$$\mathbb{E}_{\pi,\boldsymbol{I}\sim\boldsymbol{F}}[V^{\pi}(\boldsymbol{I})] \leq V(1) := 1 + \exp(-\lambda) \cdot r_2 + (1 - \exp(-\lambda)) \cdot \max\{r_1, r_2\} + O(\epsilon)$$

(iii). If π rejects all the first two queries, then conditioning on there are more than one queries with reward r_1 arriving during the interval $[0, 1]$, the optimal way for π is to serve both queries with reward r_1 if $r_1 \geq r_2$ and only serve one query with reward r_1 if $r_1 < r_2$. Then, it holds that

$$\mathbb{E}_{\pi,\boldsymbol{I}\sim\boldsymbol{F}}[V^{\pi}(\boldsymbol{I})] \leq V(2) := \exp(-\lambda)\cdot r_2 + \lambda\cdot\exp(-\lambda)\cdot(r_1+r_2) + (1-(\lambda+1)\cdot\exp(-\lambda)\cdot(r_1+\max\{r_1, r_2\})$$

Thus, we conclude that for any online algorithm π, it holds that

$$\frac{\mathbb{E}_{\pi,\boldsymbol{I}\sim\boldsymbol{F}}[V^{\pi}(\boldsymbol{I})]}{\mathbb{E}_{\boldsymbol{I}\sim\boldsymbol{F}}[V^{\text{off}}(\boldsymbol{I})]} \leq g(r_1, r_2, \lambda) := \frac{\max\{V(1), V(2), 2\}}{\hat{V}}$$

where we can neglect the $O(\epsilon)$ term by letting $\epsilon \to 0$. In this way, we can focus on the following optimization problem

$$\inf_{r_1>1, r_2>1, \lambda} g(r_1, r_2, \lambda)$$

to obtain the upper bound of the guarantee of any online algorithm relative to the prophet's value. We can numerically solve the above problem and show that when $r_1 = r_2 = 1.4119$, $\lambda = 1.2319$, the value of $g(r_1, r_2, \lambda)$ reaches its minimum and equals 0.6269, which completes our proof. □

Proof of Theorem 2.7:. For each $\boldsymbol{p} = (p_1, \ldots, p_T)$ satisfying $\sum_{t=1}^{T} p_t = k$, since each irrational number can be arbitrarily approximated by a rational number, we assume without loss of generality that p_t is a rational number for each t, i.e., $p_t = \frac{n_t}{N}$ where n_t is an integer for each t and N is an integer to denote the common denominator. We first split p_1 into $\frac{1}{N}$ and $\frac{n_1-1}{N}$ to form a new sequence $\tilde{\boldsymbol{p}} = (\frac{1}{N}, \frac{n_1-1}{N}, \frac{n_2}{N}, \ldots, \frac{n_T}{N})$. From Lemma 2.4.1, we know $\mathrm{Dual}(\boldsymbol{p}, k) \geq \mathrm{Dual}(\tilde{\boldsymbol{p}}, k)$. We then split $\frac{n_1-1}{N}$ into $\frac{1}{N}$ and $\frac{n_1-2}{N}$ and so on. In this way, we split p_1 into n_1 copies of $\frac{1}{N}$ to form a new sequence $\tilde{\boldsymbol{p}} = (\frac{1}{N}, \ldots, \frac{1}{N}, \frac{n_2}{N}, \ldots, \frac{n_T}{N})$ and Lemma 2.4.1 guarantees that $\mathrm{LP}(\boldsymbol{p}, k) \geq \mathrm{LP}(\tilde{\boldsymbol{p}}, k)$. We repeat the above operation for each t. Finally, we form a new sequence of arrival probabilities, denoted as $\boldsymbol{p}^N = (\frac{1}{N}, \ldots, \frac{1}{N}) \in \mathbb{R}^{Nk}$, and we have $\mathrm{Dual}(\boldsymbol{p}, k) \geq \mathrm{Dual}(\boldsymbol{p}^N, k)$.

From the above argument, we know that for each $\boldsymbol{p} = (p_1, \ldots, p_T)$ satisfying $\sum_{t=1}^{T} p_t = k$, there exists an integer N such that $\mathrm{Dual}(\boldsymbol{p}, k) \geq \mathrm{Dual}(\boldsymbol{p}^N, k)$, which implies that

$$\inf_{\boldsymbol{p}:\sum_t p_t = k} \mathrm{Dual}(\boldsymbol{p}, k) = \liminf_{N\to\infty} \mathrm{Dual}(\boldsymbol{p}^N, k)$$

Thus, it is enough to consider $\liminf_{N\to\infty} \mathrm{Dual}(\boldsymbol{p}^N, k)$.

We denote $\tilde{\boldsymbol{y}}_\theta(t) = (\tilde{y}_{1,\theta}(t), \ldots, \tilde{y}_{k,\theta}(t))$. We define a function $\boldsymbol{f}_\theta(\cdot) = (f_{1,\theta}(\cdot), \ldots, f_{k,\theta}(\cdot))$, where we denote $\tilde{y}_{0,\theta}(t) = 1$ and for each $l = 1, \ldots, k-1$

$$f_{l,\theta}(\tilde{y}_{1,\theta}, \ldots, \tilde{y}_{k,\theta}, t) = \begin{cases} 0, & \text{if } \tilde{y}_{l-1,\theta}(t) \leq 1-\theta \\ \tilde{y}_{l-1,\theta}(t) - (1-\theta), & \text{if } \tilde{y}_{l,\theta}(t) \leq 1-\theta \leq \tilde{y}_{l-1,\theta}(t) \\ \tilde{y}_{l-1,\theta}(t) - \tilde{y}_{l,\theta}(t), & \text{if } \tilde{y}_{l,\theta}(t) \geq 1-\theta \end{cases}$$

and

$$f_{k,\theta}(\tilde{y}_{1,\theta}, \ldots, \tilde{y}_{k,\theta}, t) = \begin{cases} 0, & \text{if } \tilde{y}_{k-1,\theta}(t) \leq 1-\theta \\ \tilde{y}_{k-1,\theta}(t) - (1-\theta), & \text{if } \tilde{y}_{k-1,\theta}(t) \geq 1-\theta \end{cases}$$

Moreover, variable $(\tilde{y}_1, \ldots, \tilde{y}_k, t)$ belongs to the feasible set of the function $f_{l,\theta}(\cdot)$ if and only if $y_{v-1} \geq y_v$ for $v = 1, \ldots, k-1$. Then, for each $\theta \in [0, 1]$, the function $\tilde{\boldsymbol{y}}_\theta(t)$ in Definition 2.6 should

be the solution to the following ordinary differential equation (ODE):

$$\frac{d\tilde{\boldsymbol{y}}_\theta(t)}{dt} = \boldsymbol{f}_\theta(\tilde{\boldsymbol{y}}_\theta, t) \text{ for } t \in [0,k] \text{ with starting point } \tilde{\boldsymbol{y}}_\theta(0) = (0,\ldots,0) \tag{A.18}$$

For each integer N and $\boldsymbol{p}^N = (\frac{1}{N},\ldots,\frac{1}{N})$ where $\|\boldsymbol{p}^N\|_1 = k$, for any fixed $\theta \in [0,1]$, we denote $\{x_{l,t}(\theta,N)\}$ as the variables constructed in Definition 2.3 under the arrival probabilities $\boldsymbol{p}^N$, where $l = 1,\ldots,k$ and $t = 1,\ldots,Nk$. We further denote $y_{l,\theta,N}(\frac{t}{N}) = \sum_{\tau=1}^{t} x_{l,\tau}(\theta,N)$ and denote $\boldsymbol{y}_{\theta,N}(\cdot) = (y_{1,\theta,N}(\cdot),\ldots,y_{k,\theta,N}(\cdot))$. It is direct to check that for each $t = 1,\ldots,Nk$, it holds that

$$(\boldsymbol{y}_{\theta,N}(\frac{t}{N}) - \boldsymbol{y}_{\theta,N}(\frac{t-1}{N}))/(\frac{1}{N}) = \boldsymbol{f}_\theta(\boldsymbol{y}_{\theta,N}, \frac{t-1}{N})$$

Thus, $\{\boldsymbol{y}_{\theta,N}(t)\}_{\forall t\in[0,k]}$ can be viewed as the result obtained from applying Euler's method (Butcher and Goodwin, 2008) to solve ODE (A.18), where there are Nk discrete points uniformly distributed within $[0,k]$. Note that for each $\theta \in [0,1]$, the function $\boldsymbol{f}_\theta(\cdot)$ is Lipschitz continuous with a Lipschitz constant 2 under infinity norm. Moreover, it is direct to note that for each $\theta \in [0,1]$ and each $t \in [0,k]$, it holds that $\|\boldsymbol{f}_\theta(\tilde{\boldsymbol{y}}, t)\|_\infty \le 1$. Then, for each $\theta \in [0,1]$, each $t_1, t_2 \in [0,k]$ and each $l = 1,\ldots,k$, we have

$$|\frac{d\tilde{y}_{l,\theta}(t_1)}{dt} - \frac{d\tilde{y}_{l,\theta}(t_2)}{dt}| \le 2 \cdot \|\tilde{\boldsymbol{y}}_\theta(t_1) - \tilde{\boldsymbol{y}}_\theta(t_2)\|_\infty \le 2 \cdot |t_1 - t_2|$$

Thus, we know that

$$|\tilde{y}_{l,\theta}(t_1) - \tilde{y}_{l,\theta}(t_2) - \frac{d\tilde{y}_{l,\theta}(t_2)}{dt} \cdot (t_1 - t_2)| \le 2 \cdot (t_1 - t_2)^2$$

We can apply the global truncation error of Euler's method (Theorem 212A (Butcher and Good-

win, 2008)) to show that $\boldsymbol{y}_{\theta,N}(k)$ converges to $\tilde{\boldsymbol{y}}_\theta(k)$ when $N \to \infty$. Specifically, we have

$$\|\boldsymbol{y}_{\theta,N}(k) - \tilde{\boldsymbol{y}}_\theta(k)\|_\infty \leq (\exp(2k) - 1) \cdot \frac{1}{N}, \quad \forall \theta \in [0, 1] \tag{A.19}$$

Now we define $Y(\theta) = \tilde{y}_{k,\theta}(k)$ as a function of $\theta \in [0, 1]$ and for each N, we define $Y_N(\theta) = y_{k,\theta,N}(k)$ as a function of $\theta \in [0, 1]$. (A.19) implies that the function sequence $\{Y_N\}_{\forall N}$ converges uniformly to the function Y when $N \to \infty$. Note that for each N, the function $Y_N(\theta)$ is continuously monotone increasing with θ due to Lemma A.0.2, then from uniform limit theorem, $Y(\theta)$ must be a continuously monotone increasing function over θ. Thus, the equation $Y(\theta) = 1-\theta$ has a unique solution, denoted as γ_k^*. For each N, we denote θ_N^* as the unique solution to the equation $Y_N(\theta) = 1 - \theta$, where we have that $\theta_N^* = \text{Dual}(\boldsymbol{p}^N, k)$. Since $\{Y_N\}_{\forall N}$ converges uniformly to the function Y, it must hold that $\gamma_k^* = \lim_{N\to\infty} \theta_N^*$, which completes our proof. □

Proof of Proposition 2.8:. The proof is the same as the proof of Proposition 3.1 in Jiang et al. (2022a).

Consider a problem setup $\hat{\mathcal{H}}$ with 4 queries and

$$(\hat{r}_1, p_1, d_1) = (r, 1, \epsilon), \quad (\hat{r}_2, p_2, d_2) = (\hat{r}_3, p_3, d_3) = (r, \frac{1-2\epsilon}{1+2\epsilon}, \frac{1}{2} + \epsilon), \quad (\hat{r}_4, p_4, d_4) = (r/\epsilon, \epsilon, 1)$$

for $r > 0$ and some $\epsilon > 0$. Obviously, if the policy π only serves queries with a size greater than $1/2$, then the expected total reward is $V_L^\pi = r + O(\epsilon)$. If the policy π only serves queries with a size no greater than $1/2$, then the expected total reward is $V_S^\pi = r$. Thus, the expected total reward of the policy π is

$$V^\pi = \max\{V_L^\pi, V_S^\pi\} = r + O(\epsilon)$$

Moreover, it is direct to see that $\sum_{t=1}^4 p_t \cdot d_t = 1$, then, we have $\text{UP}(\hat{\mathcal{H}}) = 4r$. Thus, the guarantee of π is upper bounded by $1/4 + O(\epsilon)$, which converges to $1/4$ as $\epsilon \to 0$. □

Proof of Theorem 2.9:. It is enough to prove that the threshold policy π_γ is feasible when $\gamma = \frac{1}{3+e^{-2}}$. In the remaining proof, we set $\gamma = \frac{1}{3+e^{-2}}$. For a fixed t, and any a and b, denote $\mu_{t,\gamma}(a,b] = P(a < \tilde{X}_{t,\gamma} \le b)$ assuming $\tilde{X}_{t,\gamma}$ is well-defined, it is enough to prove that $\mu_{t,\gamma}(0,1] \le 1-\gamma$ thus the random variable $\tilde{X}_{t+1,\gamma}$ is well-defined.

We define $U_t(s) = \mu_{t,\gamma}(0,s]$ for any $s \in (0,1]$. Note that by definition, we have $\mathbb{E}[\tilde{X}_{t,\gamma}] = \gamma \cdot \sum_{\tau=1}^{t} p_\tau \cdot d_\tau \le \gamma$. From integration by parts, we have that

$$\gamma \ge \mathbb{E}[\tilde{X}_{t,\gamma}] = \int_{s=0}^{1} s dU_t(s) = U_t(1) - \int_{s=0}^{1} U_t(s) ds \tag{A.20}$$

We then bound the term $\int_{s=0}^{1} U_t(s)ds$. Now suppose $U_t(1) \ge \gamma$, otherwise $U_t(1) < \gamma$ immediately implies that $U_t(1) \le 1-\gamma$, which proves our result. Then there must exists a constant $u^* \in (0,1)$ such that $\gamma \cdot u^* - \gamma \cdot \ln(u^*) = U_t(1)$. We further define

$$s^* = \begin{cases} \min\{s \in (0,1/2] : U_t(s) \ge \gamma \cdot u^*\}, & \text{if } U_t(\frac{1}{2}) \ge \gamma \cdot u^* \\ \frac{1}{2}, & \text{if } U_t(\frac{1}{2}) < \gamma \cdot u^* \end{cases}$$

Denote $U_t(s^*-) = \lim_{s \to s^*} U_t(s)$, it holds that

$$\begin{aligned} \int_{s=0}^{1} U_t(s)ds &= \int_{s=0}^{s^*-} U_t(s)ds + \int_{s=s^*}^{1/2} U_t(s)ds + \int_{s=1/2}^{1-s^*} U_t(s)ds + \int_{s=1-s^*}^{1} U_t(s)ds \\ &\le s^* \cdot (U_t(s^*-) + U_t(1)) + \int_{s=s^*}^{1/2} U_t(s)ds + \int_{s=1/2}^{1-s^*} U_t(s)ds \\ &\le s^* \cdot (2\gamma u^* - \gamma \cdot \ln(u^*)) + \int_{s=s^*}^{1/2} (2U_t(s) - \gamma \cdot \ln(\frac{U_t(s)}{\gamma}))ds \end{aligned}$$

where the last inequality holds by noting that $U_t(s^*-) \le \gamma u^*$ and for any $s \in [s^*, 1/2]$, from Lemma 2.8.1, we have $\frac{U_t(s)}{\gamma} \le \exp(-\frac{U_t(1-s)-U_t(s)}{\gamma})$, which implies that $\frac{U_t(1-s)}{\gamma} \le \frac{U_t(s)}{\gamma} - \ln(\frac{U_t(s)}{\gamma})$. Note that for any $s \in [s^*, 1/2]$, we have that $\gamma \cdot u^* \le U_t(s^*) \le U_t(s) \le U_t(1/2) \le \gamma$, where $U_t(1/2) \le \gamma$ holds directly from Lemma 2.8.1. Further note that the function $2x - \gamma \cdot \ln(x/\gamma)$ is a

convex function, thus is quasi convex. Then, for any $s \in [s^*, 1/2]$, it holds that

$$2U_t(s) - \gamma \cdot \ln(\frac{U_t(s)}{\gamma}) \leq \max\{2\gamma u^* - \gamma \cdot \ln(u^*), 2\gamma\}$$

Thus, we have that

$$\int_{s=0}^{1} U_t(s)ds \leq s^* \cdot (2\gamma u^* - \gamma \cdot \ln(u^*)) + (1/2 - s^*) \cdot \max\{2\gamma u^* - \gamma \cdot \ln(u^*), 2\gamma\}$$

If $2\gamma u^* - \gamma \cdot \ln(u^*) \leq 2\gamma$, we have $\int_{s=0}^{1} U_t(s)ds \leq 2s^*\gamma + \gamma - 2s^*\gamma = \gamma$. From (A.20), we have that $U_t(1) \leq 2\gamma < 1 - \gamma$.

If $2\gamma u^* - \gamma \cdot \ln(u^*) > 2\gamma$, we have $\int_{s=0}^{1} U_t(s)ds \leq \gamma u^* - \frac{\gamma}{2} \cdot \ln(u^*)$. From (A.20) and the definition of u^*, we have that

$$U_t(1) = \gamma u^* - \gamma \cdot \ln(u^*) \leq \gamma + \gamma u^* - \frac{\gamma}{2} \cdot \ln(u^*)$$

which implies that $u^* \geq \exp(-2)$. Note that the function $x - \ln(x)$ is non-increasing on $(0, 1)$, we have $U_t(1) \leq \gamma \cdot \exp(-2) + 2\gamma = 1 - \gamma$, which completes our proof. □

Proof of Theorem 4.11:. It is enough to consider a special case of our problem where the size of each query is deterministic. Then the proof is the same as the proof of Theorem 3.1 in Jiang et al. (2022a).

We denote by d_t the size of query t. Then, we have $\boldsymbol{p} = (p_1, \ldots, p_t)$ and $\boldsymbol{D} = (d_1, \ldots, d_T)$. Due to Lemma 2.0.2, it is enough to construct a $\boldsymbol{p}$ and $\boldsymbol{D}$ satisfying $\sum_{t=1}^{T} p_t \cdot d_t \leq 1$, such that $\text{Dual}(\boldsymbol{p}, \boldsymbol{D}) \leq \frac{1}{3+e^{-2}}$. For each $\epsilon > 0$, we consider the following $\boldsymbol{p}$ and $\boldsymbol{D}$:

$$(p_1, d_1) = (1, \epsilon), \quad (p_t, d_t) = (\frac{1 - 2\epsilon}{(T-2)(\frac{1}{2} + \epsilon)}, \frac{1}{2} + \epsilon) \text{ for all } 2 \leq t \leq T-1 \text{ and } (p_T, d_T) = (\epsilon, 1)$$

It is direct to check that $\sum_{t=1}^{T} p_t \cdot d_t = 1$. Now denote by $\{\theta^*, \alpha_t^*(c_t)\}$ the optimal solution to $\text{Dual}(\boldsymbol{p}, \boldsymbol{D})$ (we omit d_t in the expression $\alpha(d_t, c_t)$ since now d_t takes a single value). We further

denote $\mu_t = \sum_{c_t: d_t \le c_t < 1} \alpha_t^*(c_t)$ for each $2 \le t \le T-1$. Then, constraint (2.4a) implies that

$$\alpha_t^*(1) \ge \theta^* \cdot p_t - \mu_t, \quad \forall 2 \le t \le T-1$$

Thus, we have that

$$\theta^* \cdot p_T \le \alpha_T^*(1) \le p_T \cdot (1 - \alpha_1^*(1) - \sum_{t=2}^{T-1} \alpha_t^*(1)) \Rightarrow \theta^* \cdot (1 + \sum_{t=2}^{T-1} p_t) \le 1 - \alpha_1^*(1) + \sum_{t=2}^{T-1} \mu_t$$

We then consider the term $\sum_{t=2}^{T-1} \mu_t$ to upper bound θ^*. Note that since $d_t > 1/2$ for all $2 \le t \le T-1$, constraint (2.4b) implies that

$$\alpha_t^*(c_t) \le p_t \cdot (\alpha_1^*(c_t + d_1) - \sum_{\tau=2}^{t-1} \alpha_\tau^*(c_t)), \quad \forall d_t \le c_t < 1$$

Further note that $\alpha_1^*(c) = 0$ when $c < 1$, we must have that $\alpha_t^*(c_t) = 0$ when $c_t \ne 1 - d_1$. Then, we have $\mu_t = \alpha_t^*(1 - d_1)$ and

$$\mu_t \le p_t \cdot (\alpha_1^*(1) - \sum_{\tau=2}^{t-1} \mu_\tau), \quad \forall 2 \le t \le T-1$$

We then inductively show that for any $2 \le t \le T-1$, we have

$$\sum_{\tau=2}^{t} \mu_t \le (1 - \prod_{\tau=2}^{t} (1 - p_\tau)) \cdot \alpha_1^*(1) \tag{A.21}$$

When $t = 2$, (A.21) holds obviously. Now suppose that (A.21) holds for $t-1$, we have

$$\begin{aligned}\sum_{\tau=2}^{t} \mu_t &= \sum_{\tau=2}^{t-1} \mu_\tau + \mu_t \le \alpha_1^*(1) \cdot p_t + (1 - p_t) \cdot \sum_{\tau=2}^{t-1} \mu_\tau \le \alpha_1^*(1) \cdot p_t + (1 - p_t) \cdot (1 - \prod_{\tau=2}^{t-1} (1 - p_\tau)) \cdot \alpha_1^*(1) \\ &= \alpha_1^*(1) \cdot (1 - \prod_{\tau=2}^{t} (1 - p_\tau))\end{aligned}$$

Thus, (A.21) holds for all $2 \le t \le T-1$ and we have

$$\sum_{t=2}^{T-1} \mu_t \le \alpha_1^*(1)\cdot(1-\prod_{t=2}^{T-1}(1-p_t)) = \alpha_1^*(1)\cdot(1-(1-\frac{1-2\epsilon}{(T-2)(\frac{1}{2}+\epsilon)})^{T-2}) \le \alpha_1^*(1)\cdot(1-\exp(-2))+O(\epsilon)$$

which implies that

$$3\theta^* + O(\epsilon) = \theta^* \cdot (1 + \sum_{t=2}^{T-1} p_t) \le 1 - \alpha_1^*(1) + \sum_{t=2}^{T-1} \mu_t \le 1 - \alpha_1^*(1) \cdot \exp(-2) + O(\epsilon)$$

Moreover, note that constraint (2.4a) requires $\alpha_1^*(1) \ge \theta^*$, we have that $\theta^* \le \frac{1}{3+e^{-2}} + O(\epsilon)$. The proof is finished by taking $\epsilon \to 0$. □

Proof of Lemma 2.10.1:. We prove the result by induction on t. When $t = 0$, since $\mu_{0,\gamma}(0, b] = 0$ for any $0 < b \le 1/2$, the result holds trivially. Now suppose that the equation holds for $t-1$, we consider the case for t. Denote $\mathcal{F}_t$ as the support of $\tilde{d}_t$ and for each $d_t \in \mathcal{F}_t$, we denote $\eta_{t,\boldsymbol{\gamma}}(d_t)$ as the threshold defined in (2.8). Then we define the following division of $\mathcal{F}_t$:

$$\mathcal{F}_{t,1} := \{d_t \in \mathcal{F}_t : \eta_{t,\boldsymbol{\gamma}}(d_t) = 0 \text{ and } d_t \le b\}$$

$$\mathcal{F}_{t,2} = \{d_t \in \mathcal{F}_t : \eta_{t,\boldsymbol{\gamma}}(d_t) = 0 \text{ and } b < d_t \le 1-b\}$$

$$\mathcal{F}_{t,3} = \{d_t \in \mathcal{F}_t : \eta_{t,\boldsymbol{\gamma}}(d_t) > 0 \text{ and } d_t \le 1-b\}$$

$$\mathcal{F}_{t,4} = \{d_t \in \mathcal{F}_t : \eta_{t,\boldsymbol{\gamma}}(d_t) = 0 \text{ and } 1-b < d_t\}$$

$$\mathcal{F}_{t,5} = \{d_t \in \mathcal{F}_t : \eta_{t,\boldsymbol{\gamma}}(d_t) > 0 \text{ and } 1-b < d_t\}$$

Note that for each $d_t \in \mathcal{F}_t$, $\eta_{t,\boldsymbol{\gamma}}(d_t) = 0$ implies that a measure $p(d_t) \cdot (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, 1-d_t])$ of empty sample paths will be moved to d_t due to the inclusion of realization d_t when defining $\tilde{X}_{t,\boldsymbol{\gamma}}$. More specifically, the movement of sample paths due to the inclusion of each realization $d_t \in \mathcal{F}_t$ can be described as follows:

(i). For each $d_t \in \mathcal{F}_{t,1}$, obviously, $p(d_t) \cdot (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, 1-d_t])$ measure of sample paths, which is

upper bounded by $p(d_t) \cdot (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, 1-b])$, will be moved from 0 to the range $(0, b]$, while $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths will be moved out of the range $(0, b]$. Moreover, at most $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths will be moved into the range $(b, 1-b]$.

(ii). For each $d_t \in \mathcal{F}_{t,2}$, $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths will be moved out of the range $(0, b]$. Moreover, $p(d_t) \cdot (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, 1-d_t])$ measure of sample paths, which is upper bounded by $p(d_t) \cdot (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, b])$, will be moved from 0 into the range $(b, 1-b]$, while at most $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths will be moved from $(0, b]$ into $(b, 1-b]$. Thus, the measure of new sample path that is moved into the range $(b, 1-b]$ is upper bounded by $\gamma_t \cdot p(d_t)$.

(iii). For each $d_t \in \mathcal{F}_{t,3}$, then at most $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths is moved out of the range $(0, b]$, and at most $p(d_t) \cdot \mu_{t-1,\boldsymbol{\gamma}}(0, b]$ measure of sample paths is moved into the range $(b, 1-b]$.

(iv). For each $d_t \in \mathcal{F}_{t,4}$ or $d_t \in \mathcal{F}_{t,5}$, since $d_t > 1-b$, obviously, no new sample path will be added to the range $(b, 1-b]$ due to the inclusion of such realization d_t when defining $\tilde{X}_{t,\boldsymbol{\gamma}}$, while the measure of the sample paths within the range $(0, b]$ can only become smaller.

To conclude, denoting

$$\hat{p}_1 = \sum_{d_t \in \mathcal{F}_{t,1}} p(d_t) \text{ and } \hat{p}_2 = \sum_{d_t \in \mathcal{F}_{t,2}} p(d_t) \text{ and } \hat{p}_3 = \sum_{d_t \in \mathcal{F}_{t,3}} p(d_t)$$

we have that

$$\mu_{t,\boldsymbol{\gamma}}(0, b] \leq \mu_{t-1,\boldsymbol{\gamma}}(0, b] + (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0, 1-b] - \mu_{t-1,\boldsymbol{\gamma}}(0, b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0, b] \cdot \hat{p}_2 - \mu_{t-1,\boldsymbol{\gamma}}(0, b] \cdot \hat{p}_3 \tag{A.22}$$

and

$$\mu_{t,\boldsymbol{\gamma}}(b, 1-b] \leq \mu_{t-1,\boldsymbol{\gamma}}(b, 1-b] + \mu_{t-1,\boldsymbol{\gamma}}(0, b] \cdot \hat{p}_1 + \gamma_t \cdot \hat{p}_2 + \mu_{t-1,\boldsymbol{\gamma}}(0, b] \cdot \hat{p}_3 \tag{A.23}$$

Moreover, it holds that $\hat{p}_1 + \hat{p}_2 + \hat{p}_3 \leq 1$. We now consider the following two cases separately.

Case 1: If $\hat{p}_1 > 0$, then we must have $\gamma_t \geq \mu_{t-1,\boldsymbol{\gamma}}(0, 1-b]$. Notice that $\hat{p}_1 \leq 1 - \hat{p}_2$, from (A.22), we have

$$\begin{aligned}
\mu_{t,\boldsymbol{\gamma}}(0,b] &\leq \mu_{t-1,\boldsymbol{\gamma}}(0,b] + (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0,1-b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0,b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_2 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3 \\
&\leq \mu_{t-1,\boldsymbol{\gamma}}(0,b] + (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(0,1-b]) \cdot (1-\hat{p}_2) - \mu_{t-1,\boldsymbol{\gamma}}(0,b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_2 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3 \\
&= (\gamma_t - \mu_{t-1,\boldsymbol{\gamma}}(b,1-b]) \cdot (1-\hat{p}_2) - \mu_{t-1,\boldsymbol{\gamma}}(0,b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3 \\
&\leq (\gamma_1 - \mu_{t-1,\boldsymbol{\gamma}}(b,1-b]) \cdot (1-\hat{p}_2) - \mu_{t-1,\boldsymbol{\gamma}}(0,b]) \cdot \hat{p}_1 - \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3 \qquad \text{(A.24)}
\end{aligned}$$

where the last inequality holds from $\gamma_1 \geq \gamma_t$. Moreover, from (A.23), we have that

$$\begin{aligned}
\exp(-\frac{1}{\gamma_1} \cdot \mu_{t,\boldsymbol{\gamma}}(b,1-b]) &\geq \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \frac{\gamma_t \hat{p}_2}{\gamma_1}) \cdot \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3) \\
&\geq \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) \cdot \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3) \\
&\geq \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) \cdot (1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3) \\
&= \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) \\
&\quad - \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) \cdot (\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_1 + \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3) \\
&\geq \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(0,b] \cdot \hat{p}_3
\end{aligned} \qquad \text{(A.25)}$$

where the second inequality holds from $\gamma_1 \geq \gamma_t$, the third inequality holds from $\exp(-x) \geq 1 - x$ for any $x \geq 0$ and the last inequality holds from $\exp(-x) \leq 1$ for any $x \geq 0$. Further note that

$$\exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] - \hat{p}_2) = \exp(-\frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b]) \cdot \exp(-\hat{p}_2) \geq (1 - \frac{1}{\gamma_1} \cdot \mu_{t-1,\boldsymbol{\gamma}}(b,1-b]) \cdot (1-\hat{p}_2)$$

From (A.24) and (A.25), we have

$$\begin{aligned}\exp(-\frac{1}{\gamma_1}\cdot\mu_{t,\boldsymbol{\gamma}}(b,1-b]) &\geq (1-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(b,1-b])\cdot(1-\hat{p}_2)-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_1-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3\\ &\geq \frac{1}{\gamma_1}\cdot\mu_{t,\boldsymbol{\gamma}}(0,b]\end{aligned}$$

Case 2: If $\hat{p}_1 = 0$, then we have

$$\mu_{t,\boldsymbol{\gamma}}(0,b] \leq \mu_{t-1,\boldsymbol{\gamma}}(0,b] - \mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_2 - \mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3 \tag{A.26}$$

and

$$\mu_{t,\boldsymbol{\gamma}}(b,1-b] \leq \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] + \gamma_t\cdot\hat{p}_2 + \mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3 \leq \mu_{t-1,\boldsymbol{\gamma}}(b,1-b] + \gamma_1\cdot\hat{p}_2 + \mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3$$

Thus, it holds that

$$\begin{aligned}\exp(-\frac{1}{\gamma_1}\cdot\mu_{t,\boldsymbol{\gamma}}(b,1-b]) &\geq \exp(-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(b,1-b]-\hat{p}_2)\cdot\exp(-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3)\\ &\geq \exp(-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(b,1-b]-\hat{p}_2)\cdot(1-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3)\\ &\geq \exp(-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(b,1-b]-\hat{p}_2)-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3\\ &\geq \exp(-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(b,1-b])\cdot(1-\hat{p}_2)-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3\\ &\geq \frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot(1-\hat{p}_2)-\frac{1}{\gamma_1}\cdot\mu_{t-1,\boldsymbol{\gamma}}(0,b]\cdot\hat{p}_3\end{aligned} \tag{A.27}$$

where the third inequality holds from $\exp(-a) \leq 1$ for any $a \geq 0$ and the last inequality holds from induction hypothesis. Our proof is completed immediately by combining (A.26) and (A.27).

□

Proof of Theorem 2.11:. For each fixed t, we define $U_t(s) = \mu_{t,\boldsymbol{\gamma}}(0,s] = P(0 < \tilde{X}_{t,\boldsymbol{\gamma}} \leq s)$ for any $s \in (0,1]$. Note that by Algorithm 2, we have $\mathbb{E}[\tilde{X}_{t,\boldsymbol{\gamma}}] = \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau$. From integration by parts, we have that

$$\sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau = \mathbb{E}[\tilde{X}_{t,\boldsymbol{\gamma}}] = \int_{s=0}^{1} U_t(s)ds = U_t(1) - \int_{s=0}^{1} U_t(s)ds \tag{A.28}$$

We then bound the term $\int_{s=0}^{1} U_t(s)ds$. If $U_t(1) < \gamma_1$, then we immediately have

$$P(\tilde{X}_{t,\boldsymbol{\gamma}} = 0) > 1 - \gamma_1 \geq 1 - \gamma_1 - \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau$$

which proves (2.9). Thus, in the remaining part of the proof, it is enough for us to only focus on the case $U_t(1) \geq \gamma_1$.

If $U_t(1) \geq \gamma_1$, then there must exists a constant $u^* \in (0,1)$ such that

$$\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*) = U_t(1).$$

We further define

$$s^* = \begin{cases} \min\{s \in (0,1/2] : U_t(s) \geq \gamma_1 \cdot u^*\}, & \text{if } U_t(\frac{1}{2}) \geq \gamma_1 \cdot u^* \\ \frac{1}{2}, & \text{if } U_t(\frac{1}{2}) < \gamma_1 \cdot u^* \end{cases}$$

Following the proof of Theorem 2.9, we can show that

$$\int_{s=0}^{1} U_t(s)ds \leq s^* \cdot (2\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*)) + (1/2 - s^*) \cdot \max\{2\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*), 2\gamma_1\}$$

We further simplify the above expression separately by comparing the value of $2\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*)$ and $2\gamma_1$.

Case 1: If $2\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*) \leq 2\gamma_1$, we have $\int_{s=0}^{1} U_t(s)ds \leq 2s^*\gamma_1 + \gamma_1 - 2s^*\gamma_1 = \gamma_1$. From (A.28),

we have that

$$U_t(1) \leq \gamma_1 + \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau$$

Case 2: If $2\gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*) > 2\gamma_1$, we have $\int_{s=0}^{1} U_t(s)ds \leq \gamma_1 \cdot u^* - \frac{\gamma_1}{2} \cdot \ln(u^*)$. From (A.28) and the definition of u^*, we have that

$$U_t(1) = \gamma_1 \cdot u^* - \gamma_1 \cdot \ln(u^*) \leq \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau + \gamma_1 \cdot u^* - \frac{\gamma_1}{2} \cdot \ln(u^*)$$

The above inequality implies that

$$u^* \geq \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)$$

Note that the function $x - \ln(x)$ is non-increasing on $(0, 1)$, hence we have

$$U_t(1) \leq 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau + \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)$$

Combing the above two cases, we conclude that

$$U_t(1) \leq \max\{\gamma_1 + \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau, \quad 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau + \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)\}$$

Note that $P(\tilde{X}_{t,\boldsymbol{\gamma}} = 0) = 1 - U_t(1)$, we conclude that

$$P(\tilde{X}_{t,\boldsymbol{\gamma}} = 0) \geq \min\{1 - \gamma_1 - \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau, \quad 1 - 2 \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau - \gamma_1 \cdot \exp(-\frac{2}{\gamma_1} \cdot \sum_{\tau=1}^{t} \gamma_\tau \cdot \psi_\tau)\}$$

which completes our proof. □

Proof of Lemma 2.11.1:. Since the function $h_{\gamma_0}(\cdot)$ is non-increasing and non-negative over $[0,1]$, it is direct to see that

$$1 \geq \hat{\gamma}_1 \geq \cdots \geq \hat{\gamma}_T \geq 0$$

Note that for each $t = 1, \ldots, T$, we have

$$\int_{\tau=0}^{k_t} h_{\gamma_0}(\tau) d\tau = \sum_{\tau=1}^{t} \hat{\gamma}_\tau \cdot \psi_\tau$$

and $\gamma_0 \geq \hat{\gamma}_1$. Then, for each $t = 1, \ldots, T-1$ and each $\tau \in [k_t, k_{t+1}]$, it holds that

$$h_{\gamma_0}(\tau) \leq 1 - \gamma_0 - \int_{\tau'=0}^{\tau} h_{\gamma_0}(\tau') d\tau' \leq 1 - \gamma_0 - \int_{\tau'=0}^{k_t} h_{\gamma_0}(\tau') d\tau' \leq 1 - \hat{\gamma}_1 - \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'}$$

which implies that

$$\hat{\gamma}_{t+1} \leq 1 - \hat{\gamma}_1 - \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'}$$

since $\hat{\gamma}_{t+1}$ is defined as the average of function $h_{\gamma_0}(\cdot)$ over $[k_t, k_{t+1}]$ in (2.13).

Similarly, note that the function $2x + \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot x)$ is monotone increasing when $x \geq 0$. Then, for each $t = 1, \ldots, T-1$ and each $\tau \in [k_t, k_{t+1}]$, we have

$$\begin{aligned}
h_{\gamma_0}(\tau) &\leq 1 - 2 \cdot \int_{\tau'=0}^{\tau} h_{\gamma_0}(\tau') d\tau' - \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot \int_{\tau'=0}^{\tau} h_{\gamma_0}(\tau') d\tau') \\
&\leq 1 - 2 \cdot \int_{\tau'=0}^{k_t} h_{\gamma_0}(\tau') d\tau' - \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot \int_{\tau'=0}^{k_t} h_{\gamma_0}(\tau') d\tau') \\
&= 1 - 2 \cdot \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'} - \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'})
\end{aligned}$$

which implies that

$$\hat{\gamma}_{t+1} \leq 1 - 2 \cdot \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'} - \gamma_0 \cdot \exp(-\frac{2}{\gamma_0} \cdot \sum_{\tau'=1}^{t} \hat{\gamma}_{\tau'} \cdot \psi_{\tau'})$$

since $\hat{\gamma}_{t+1}$ is defined as the average of function $h_{\gamma_0}(\cdot)$ over $[k_t, k_{t+1}]$ in (2.13). Thus, we conclude that $\{\hat{\gamma}_t\}_{t=1}^{T}$ is a feasible solution to OP($\boldsymbol{\psi}$). □

B | Appendix for Chapter 3

Proof of Theorem 3.4. We denote by X_j the number of agents that have a type j and are accepted by the prophet. Clearly, we have

$$\mathsf{Proph}(I) = \sum_{j=1}^{m} r_j \cdot \mathbb{E}[X_j] = \sum_{j=1}^{m} \Delta_j \cdot \mathbb{E}[\sum_{j'=1}^{j} X_{j'}]$$

Moreover, note that the distribution of $\sum_{j'=1}^{j} X_{j'}$ is given by $\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}$. Thus, it holds that

$$\mathsf{Proph}(I) = \sum_{j=1}^{m} \Delta_j \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$$

We now denote by $\{a_j^*\}_{j=1}^m$ one optimal solution to $\mathsf{ExAnte}(I)$. Clearly, for any $j_1 < j_2$, if $a_{j_1}^* < \sum_{t=1}^{n} p_{tj_1}$, it must hold that $a_{j_2}^* = 0$. Otherwise, we construct another set of solution $\{a'_j\}_{j=1}^m$ by letting

$$a'_j = a_j^* \text{ for } j \neq j_1 \text{ and } j_2, \quad a'_{j_1} = a_{j_1}^* + \epsilon, \quad a'_{j_2} = a_{j_2}^* - \epsilon$$

where $0 < \epsilon \leq \min\{a_{j_2}^*, \sum_{t=1}^{n} p_{tj_1} - a_{j_1}^*\}$. Clearly, $\{a'_j\}_{j=1}^m$ is feasible to $\mathsf{ExAnte}(I)$ and yields a higher objective value than $\{a_j^*\}_{j=1}^m$, which contradicts the optimality of $\{a_j^*\}_{j=1}^m$. Thus, it holds that

$$\sum_{j'=1}^{j} a_{j'}^* = \min\{\sum_{t=1}^{n} G_{tj}, k\} \text{ for any } j$$

and

$$\mathsf{ExAnte}(I) = \sum_{j=1}^{m} r_j \cdot a_j^* = \sum_{j=1}^{m} \Delta_j \cdot \sum_{j'=1}^{j} a_{j'}^* = \sum_{j=1}^{m} \Delta_j \cdot \min\{\sum_{t=1}^{n} G_{tj}, k\}$$

which completes our proof of (3.1). □

Proof of Lemma 3.6.1. Given a feasible solution to (3.3), we construct the following feasible solution to (3.4) with the same value of θ. Let $y_t^l = \sum_{j=1}^{m} y_{tj}^l$ for all $t \in [T]$ and $l \in [k]$. Clearly constraints (3.4d)–(3.4e) are satisfied. Constraints (3.4c) are satisfied due to the fact that $\sum_{j=1}^{m} p_{tj} = 1$ for all t. Finally, it remains to show that $\min\{y_t^l, G_{tj} x_t^l\} \geqslant \sum_{j'=1}^{j} y_{tj'}^l$ which would make constraints (3.4b) satisfied. To see this, note that $\min\{y_t^l, G_{tj} x_t^l\} = \min\{\sum_{j=1}^{m} y_{tj}^l, x_t^l \sum_{j'=1}^{j} p_{tj'}\} \geqslant \min\{\sum_{j=1}^{m} y_{tj}^l, \sum_{j'=1}^{j} y_{tj'}^l\}$ where the inequality applies (3.3c). Since both arguments in the min are at least $\sum_{j'=1}^{j} y_{tj'}^l$, this completes the proof.

Conversely, given a feasible solution to (3.4), we construct the following feasible solution to (3.3) with the same value of θ, which is the harder direction. For each $t \in [T]$ and $l \in [k]$, we iteratively define

$$\begin{aligned} y_{t1}^l &= \min\{y_t^l, p_{t1} x_t^l\} \\ y_{t2}^l &= \min\{y_t^l - y_{t1}^l, p_{t2} x_t^l\} \\ &\dots \end{aligned}$$

$$y_{im}^l = \min\{y_t^l - \sum_{j=1}^{m-1} y_{tj}^l, p_{im} x_t^l\}.$$

Constraints (3.3c) hold from the second argument in the min, while constraints (3.3e) hold because by the first argument in the min, the sum $\sum_j y_{tj}^l$ can never exceed y_t^l. Meanwhile, it can be inductively established that $\sum_{j'=1}^{j} y_{tj'}^l = \min\{y_t^l, x_t^l \sum_{j'=1}^{j} p_{tj'}\} = \min\{y_t^l, G_{tj} x_t^l\}$ for all $j = 1, \dots, m$, establishing constraints (3.3b). Finally, by the same fact $\sum_{j=1}^{m} y_{tj}^l = \min\{y_t^l, x_t^l\} = y_t^l$, establishing

constraints (3.3d) and completing the proof. □

Proof of Theorem 3.12. We consider the dual of LP (3.5). We introduce x_t^l as the dual variable for constraint (3.5b), the dual variable y_{tj}^l for constraint (3.5c) and dual variable θ for constraint (3.5d). Then, we get the following LP as the dual of LP (3.5).

$$\max \ \theta \tag{B.1a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \sum_{j'=1}^{j} y_{tj'}^l \qquad \forall j \in [m] \tag{B.1b}$$

$$y_{tj}^l = \begin{cases} p_{tj} x_t^l, & j < J \\ p_{tJ} \rho x_t^l, & j = J \\ 0, & j > J \end{cases} \qquad \forall t \in [T], l \in [k] \tag{B.1c}$$

$$x_t^l = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l - \sum_{j=1}^{m} y_{t-1,j}^l + \sum_{j=1}^{m} y_{t-1,j}^{l+1}, & t > 1 \end{cases} \qquad \forall t \in [T], l \in [k] \tag{B.1d}$$

$$\theta, x_t^l, y_{tj}^l \in \mathbb{R} \qquad \forall t \in [T], \forall j \in [m], \forall l \in [k] \tag{B.1e}$$

For any $t \in [T], l \in [k]$, we define $y_t^l = \sum_{j=1}^{m} y_{tj}^l$. Then, constraint (B.1c) implies that

$$y_t^l = (\sum_{j<J} p_{tj} + p_{tJ}\rho) x_t^l = ((1-\rho) G_{t,J-1} + \rho G_{tJ}) x_t^l, \ \ \forall t \in [T], l \in [k].$$

Moreover, for any $j \in [m]$, constraint (B.1c) implies that

$$\sum_{j'=1}^{j} y_{tj'}^l = \min\{y_t^l, G_{tj} x_t^l\}, \ \ \forall t \in [T], l \in [k]$$

Thus, a feasible solution to LP (B.1) can be translated into a feasible solution to the following LP, with the same objective value,

$$\max \ \theta \tag{B.2a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\} \qquad \forall j \in [m] \tag{B.2b}$$

$$y_t^l = (\sum_{j<J} p_{tj} + p_{tJ}\rho) x_t^l = ((1-\rho) G_{t,J-1} + \rho G_{tJ}) x_t^l \qquad \forall t \in [T], l \in [k] \tag{B.2c}$$

$$x_t^l = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l - y_{t-1}^l + y_{t-1}^{l+1}, & t > 1 \end{cases} \qquad \forall t \in [T], l \in [k] \tag{B.2d}$$

$$y_t^l \geqslant 0 \qquad \forall t \in [T], l \in [k]. \tag{B.2e}$$

We now show that a feasible solution to LP (B.2), denoted by $\{\theta, x_t^l, y_t^l\}$, can be translated into a feasible solution to LP (B.1) with the same objective value, which implies that the objective value of LP (B.1) is equivalent to the objective value of LP (B.2). To be specific, we define

$$y_{tj}^l = \begin{cases} p_{tj} x_t^l, & j < J \\ p_{tJ} \rho x_t^l, & j = J \\ 0, & j > J \end{cases}$$

Clearly, $\{\theta, x_t^l, y_{tj}^l\}$ satisfy the constraint (B.1c) and (B.1d). We also have $\sum_{j'=1}^{j} y_{tj'}^l = \min\{y_t^l, G_{tj} x_t^l\}$ for any $t \in [T], l \in [k], j \in [m]$, which implies that constraint (B.1b) is satisfied. Thus, $\{\theta, x_t^l, y_{tj}^l\}$ is a feasible solution to LP (B.1).

For the problem instance $I = (\mathbf{G}, \mathbf{\Delta})$, we denote by $\mathrm{OST}_{J,\rho}(I)$ the total expected reward collected by the oblivious static threshold policy J, ρ on problem instance I. Then, from the definition

of LP (3.5), we have that

$$\inf_{\Delta} \frac{\mathsf{OST}_{J,\rho}(I)}{\mathsf{Proph}(I)} = \text{LP (3.5) with variable } Q_j = \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$$

Similarly, we have

$$\inf_{\Delta} \frac{\mathsf{OST}_{J,\rho}(I)}{\mathsf{ExAnte}(I)} = \text{LP (3.5) with variable } Q_j = \min\{\sum_{t=1}^{n} G_{tj}, k\}$$

Note that LP (B.1) is the dual of LP (3.5), and we have shown the objective value of LP (B.1) is equivalent to the objective value of LP (B.2), which gives the expression of $\mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G})$ (resp. $\mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G})$) when $Q_j = \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$ (resp. $Q_j = \min\{\sum_{t=1}^{n} G_{tj}, k\}$). Thus, we have

$$\inf_{\Delta} \frac{\mathsf{OST}_{J,\rho}(I)}{\mathsf{Proph}(I)} = \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G}) \text{ and } \inf_{\Delta} \frac{\mathsf{OST}_{J,\rho}(I)}{\mathsf{ExAnte}(I)} = \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G})$$

Note that according to Theorem 3.9, the choice of J, ρ for OST can depend on $\mathbf{G}$. Thus, we know that the tight guarantee for OST relative to the prophet (resp. ex-ante relaxation) is given by

$$\inf_{\mathbf{G}} \sup_{J,\rho} \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G}) \text{ (resp. } \inf_{\mathbf{G}} \sup_{J,\rho} \mathrm{innerLP}_{k,T}^{\mathsf{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G}))$$

which completes our proof. □

Proof of Theorem 3.16. We consider the dual of LP (3.7). We introduce the dual variable $x_t^l(J,\rho)$ for constraint (3.7b), the dual variable $y_{tj}^l(J,\rho)$ for constraint (3.7c), the dual variable $\mu(J,\rho)$ for constraint (3.7d) and the dual variable θ for constraint (3.7e). Then, we get the follow-

ing LP as the dual of LP (3.7).

$$\max \theta \tag{B.3a}$$

$$\text{s.t. } \theta \cdot Q_j \leqslant \int_{J,\rho} \mu(J,\rho) \left(\sum_{t=1}^{T} \sum_{l=1}^{k} \sum_{j'=1}^{j} y_{tj'}^{l}(J,\rho) \right) \quad \forall j \in [m] \tag{B.3b}$$

$$y_{tj}^{l}(J,\rho) = \begin{cases} p_{tj} x_t^l(J,\rho), & j < J \\ p_{tJ} \rho x_t^l(J,\rho), & j = J \\ 0, & j > J \end{cases} \quad \forall t \in [T], l \in [k], J \in [m], \rho \in (0,1] \tag{B.3c}$$

$$x_t^l(J,\rho) = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l(J,\rho) - \sum_{j=1}^{m} y_{t-1,j}^{l}(J,\rho) + \sum_{j=1}^{m} y_{t-1,j}^{l+1}(J,\rho), & t > 1 \end{cases} \quad \forall t \in [T], l \in [k] \tag{B.3d}$$

$$\mu(J,\rho) \geqslant 0 \quad \forall J \in [m], \rho \in (0,1] \tag{B.3e}$$

Note that we can select a positive $\{x_t^l(J,\rho), y_{tj}^l(J,\rho)\}$ satisfying constraints (B.3c) and (B.3d), select $\mu(J,\rho)$ to be a uniform distribution over $J \in [m], \rho \in (0,1]$, and set $\theta = 0$. Then, all the inequality constraints in LP (B.3) can be satisfied as strict inequalities by $\{\theta, x_t^l(J,\rho), y_{tj}^l(J,\rho), \mu(J,\rho)\}$. Thus, the Slater's condition is satisfied and strong duality hols between LP (3.7) and LP (B.3) (Theorem 2.3 in Shapiro (2009)).

Now we define $y_t^l(J,\rho) = \sum_{j=1}^{m} y_{tj}^l(J,\rho)$ for any $t \in [T], l \in [k], J \in [m], \rho \in (0,1]$. Then, constraint (B.3c) implies that

$$y_t^l(J,\rho) = ((1-\rho)G_{t,J-1} + \rho G_{tJ}) x_t^l(J,\rho), \quad \forall t \in [T], l \in [k], J \in [m], \rho \in (0,1]$$

and

$$\sum_{j'=1}^{j} y_{tj'}^{l}(J,\rho) = \min\{y_t^l(J,\rho), G_{tj}x_t^l(J,\rho)\}, \quad \forall t \in [T], l \in [k], j, J \in [m], \rho \in (0,1]$$

Thus, a feasible solution to LP (B.3) can be translated into a feasible solution to LP (3.8) with the same objective value.

On the other hand, we denote by $\{\theta, \mathbf{x}(J,\rho), \mathbf{y}(J,\rho), \mu(J,\rho)\}$ a feasible solution to LP (3.8). Then, for any $t \in [T], l \in [k], J \in [m], \rho \in (0,1]$, we define

$$y_{tj}^{l}(J,\rho) = \begin{cases} p_{tj}x_t^l(J,\rho), & j < J \\ p_{tJ}\rho x_t^l(J,\rho), & j = J \\ 0, & j > J \end{cases}$$

Clearly, we have $\sum_{j'=1}^{j} y_{tj'}^{l}(J,\rho) = \min\{y_t^l(J,\rho), G_{tj}x_t^l(J,\rho)\}$ for any $t \in [T], l \in [k], j, J \in [m], \rho \in (0,1]$ and constraint (3.8c) implies that $\sum_{j=1}^{m} y_{tj}^{l}(J,\rho) = y_t^l(J,\rho)$ for any $t \in [T], l \in [k], J \in [m], \rho \in (0,1]$. Then, we have $\{\theta, x_t^l(J,\rho), y_{tj}^{l}(J,\rho), \mu(J,\rho)\}$ a feasible solution to LP (B.3) with the same objective value. Thus, the objective value of LP (B.3) is equivalent to the objective value of LP (3.8).

For the problem instance $I = (\mathbf{G}, \mathbf{\Delta})$, we denote by $\mathsf{ST}_{J,\rho}(I)$ the total expected reward collected by the static threshold policy J, ρ on problem instance I. Then, from the definition of LP (3.7), we have that

$$\inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{Proph}(I)} = \text{LP (3.7) with variable } Q_j = \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$$

Similarly, we have

$$\inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{ExAnte}(I)} = \text{LP (3.7) with variable } Q_j = \min\{\sum_{t=1}^{n} G_{tj}, k\}$$

Note that LP (B.3) is the dual of LP (3.7), where strong duality holds, and we have shown the objective value of LP (B.3) is equivalent to the objective value of LP (3.8), which gives the expression of $\mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{Proph}}(\mathbf{G})$ (resp. $\mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G})$) when $Q_j = \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G_{tj}), k\}]$ (resp. $Q_j = \min\{\sum_{t=1}^{n} G_{tj}, k\}$). Thus, we have

$$\inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{Proph}(I)} = \mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{Proph}}(\mathbf{G}) \text{ and } \inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{ExAnte}(I)} = \mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G})$$

Note that according to Theorem 3.9, the choice of J, ρ for ST can depend both on $\mathbf{G}$ and $\mathbf{\Delta}$. Thus, we know that the tight guarantee for ST relative to the prophet is given by

$$\inf_{\mathbf{G}} \inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{Proph}(I)} = \inf_{\mathbf{G}} \mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{Proph}}(\mathbf{G})$$

and the tight guarantee for ST relative to the ex-ante relaxation is given by

$$\inf_{\mathbf{G}} \inf_{\Delta} \sup_{J,\rho} \frac{\mathsf{ST}_{J,\rho}(I)}{\mathsf{ExAnte}(I)} = \inf_{\mathbf{G}} \mathrm{innerLP}_{k,T}^{\mathsf{ST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G}).$$

which completes our proof. □

Proof of Theorem 3.17. First we show that $\inf_{\mathbf{G}} \mathrm{innerLP}_{k,T}^{\mathsf{DP}/\mathsf{ExAnte}}(\mathbf{G}) \geqslant$ (3.11). Given a $\mathbf{G}$ for the LHS, we construct an instance defined by $g_1, \ldots, g_T$ for the inner maximization problem on the RHS and show that its optimal solution forms a feasible solution to $\mathrm{innerLP}_{k,T}^{\mathsf{DP}/\mathsf{ExAnte}}(\mathbf{G})$, which would be sufficient. To accomplish this, let $J \in [m]$, $\rho \in (0, 1]$ be such that $\sum_{t=1}^{T}(G_{t,J-1} + p_{tJ}\rho) = k$, which must uniquely exist since $G_{im} = 1$ for all $t \in [T]$ and $T > k$. Define $g_t = G_{t,J-1} + p_{tJ}\rho$ for all t, which we take to be our instance for the RHS, and consider an optimal solution defined by

$\theta, \mathbf{x}, \mathbf{y}$ for its inner problem. We claim that this forms a feasible solution to $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$. To see why, note that for any t, the expression

$$\frac{\sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\}}{G_{tj}} = \sum_{l=1}^{k} \min\{\frac{y_t^l}{G_{tj}}, x_t^l\}$$

is decreasing in G_{tj}. Therefore, for all $j < J$, since $G_{tj} \leqslant g_t$, we have

$$\frac{\sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\}}{G_{tj}} \geqslant \frac{\sum_{l=1}^{k} \min\{y_t^l, g_i x_t^l\}}{g_t} \geqslant \theta.$$

where the final inequality applies (3.11b). Therefore, for all $j < J$, we deduce

$$\frac{\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\}}{\min\{\sum_{t=1}^{n} G_{tj}, k\}} = \frac{\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\}}{\sum_{t=1}^{n} G_{tj}} \geqslant \theta$$

as required for (3.10b). Meanwhile, for all $j \geqslant J$, since $G_{tj} \geqslant g_t$, we directly have

$$\frac{\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, G_{tj} x_t^l\}}{\min\{\sum_{t=1}^{n} G_{tj}, k\}} \geqslant \frac{\sum_{t=1}^{T} \sum_{l=1}^{k} \min\{y_t^l, g_i x_t^l\}}{\sum_{t=1}^{T} g_t} \geqslant \theta$$

which completes the proof that $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G}) \geqslant$ (3.11).

To show that $\inf_{\mathbf{G}} \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G}) \leqslant$ (3.11), which is the harder direction, given an instance defined by $g_1, \ldots, g_T$ for the outer problem in (3.11) such that $\sum_{t=1}^{n} g_t \leq k$, we construct an instance $\mathbf{G}$ for which $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\mathbf{G})$ is no greater than the optimal objective value of the inner problem in (3.11).

The construction goes as follows. Define $m = T + 1$. We let

$$G_{tj} = g_t \cdot \mathbb{1}(t \leq j) \text{ for all } j < m \text{ and } G_{im} = 1, \forall t \in [T] \tag{B.4}$$

Note that the feasibility constraints of $G_{t1} \leqslant \cdots \leqslant G_{im} = 1$ are satisfied for all t. Under this

construction, it holds that

$$Q_j = \min\{\sum_{t=1}^{n} G_{tj}, k\} = \min\{\sum_{t=1}^{j} g_t, k\} = \sum_{t=1}^{j} g_t, \forall j \in [m]$$

where $g_m = k - \sum_{t=1}^{n} g_t$. Now, denote by $\{\Delta_j^*, U_{tj}^{l*}, V_t^{k*}\}$ one optimal solution of the following LP:

$$\min V_1^k \tag{B.5a}$$

$$\text{s.t. } V_t^l = \sum_{j=1}^{m} p_{tj} U_{tj}^l + V_{t+1}^l \qquad \forall t \in [T], l \in [k] \tag{B.5b}$$

$$U_{tj}^l \geqslant \sum_{j'=j}^{m} \Delta_{j'} - (V_{t+1}^l - V_{t+1}^{l-1}) \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.5c}$$

$$\sum_{j'=j}^{m} \Delta_{j'} \geqslant 0 \qquad \forall j \in [m] \tag{B.5d}$$

$$\sum_{j=1}^{m} Q_j \Delta_j = 1 \tag{B.5e}$$

$$\Delta_j \in \mathbb{R}, \Delta_m = 0, U_{tj}^l \geqslant 0 \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.5f}$$

with $p_{tj} = G_{tj} - G_{t,j-1}$. We further denote by $r_j^* = \sum_{j'=j}^{m} \Delta_{j'}^*$ for each $j \in [m]$ and denote by $\{\sigma(j), \forall j = 1, \ldots, m\}$ a permutation of $\{1, \ldots, m\}$ such that $r_{\sigma(1)}^* \geq r_{\sigma(2)}^* \geq \ldots \geq r_{\sigma(m)}^*$. For each $j \in [m]$, we further denote $\hat{G}_{tj} = \sum_{j'=1}^{j} p_{t\sigma(j')}$, for all $t \in [T]$. Then we have the following claim.

Claim B.0.1. *It holds that* LP(B.5) $\geq \text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\hat{\mathbf{G}})$.

On the other hand, the dual of LP(B.5) is given as

$$\max \theta \tag{B.6a}$$

$$\text{s.t. } \theta \cdot Q_j + \sum_{j'=1}^{j} \alpha_{j'} = \sum_{t=1}^{T} \sum_{l=1}^{k} \sum_{j'=1}^{j} y_{tj'}^{l} \qquad \forall j \in [T] \tag{B.6b}$$

$$y_{tj}^{l} \leqslant p_{tj} x_t^l \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.6c}$$

$$x_t^l = \begin{cases} 1, & t = 1, l = k \\ 0, & t = 1, l < k \\ x_{t-1}^l - \sum_{j=1}^{m} (y_{t-1,j}^{l} - y_{t-1,j}^{l+1}), & t > 1 \end{cases} \qquad \forall t \in [T], l \in [k] \tag{B.6d}$$

$$y_{tj}^{l} \geqslant 0, \alpha_j \geq 0, \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.6e}$$

Denote by $\{\theta^*, \alpha_j^*, x_t^{l*}, y_{tj}^{l*}\}$ one optimal solution of LP(B.6) and define $y_t^{l*} = \sum_{j=1}^{m} y_{tj}^{l*}$ for each $t \in [T], l \in [k]$. We now show that $\{\theta^*, x_t^{l*}, y_t^{l*}\}$ is a feasible solution to LP(3.11).

Clearly, $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{P}_T^k$. From constraint (B.6b), for $j \in [T]$, we have that

$$\theta^* \cdot (Q_j - Q_{j-1}) + \alpha_j^* = \sum_{l=1}^{k} y_{jj}^{l*} \leq \sum_{l=1}^{k} \min\{y_j^{l*}, g_j x_t^{l*}\}$$

where the last inequality follows from the construction of $\mathbf{G}$ and constraint (B.6c). Further note that for $j \in [T]$, we have $\theta^* \cdot g_j \leq \theta^* \cdot (Q_j - Q_{j-1}) + \alpha_j^*$. We conclude that $\{\theta^*, x_t^{l*}, y_t^{l*}\}$ is a feasible solution to LP(3.11). Thus, from Claim B.0.1, we have that

$$\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\hat{\mathbf{G}}) \leq \text{LP(B.5)} = \text{LP(B.6)} = \theta^* \leq \text{LP(3.11)}$$

which completes our proof. □

Proof of Claim B.0.1. Note that from Lemma 3.6.1, $\text{innerLP}_{k,T}^{\text{DP/ExAnte}}(\hat{\mathbf{G}})$ is given as the optimal objective value of the following LP:

$$\min V_1^k \tag{B.7a}$$

$$\text{s.t. } V_t^l = \sum_{j=1}^{m} \hat{p}_{tj} U_{tj}^l + V_{t+1}^l \qquad \forall t \in [T], l \in [k] \tag{B.7b}$$

$$U_{tj}^l \geqslant \sum_{j'=j}^{m} \Delta_{j'} - (V_{t+1}^l - V_{t+1}^{l-1}) \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.7c}$$

$$\sum_{j=1}^{m} \hat{Q}_j \Delta_j = 1 \tag{B.7d}$$

$$\Delta_j \geq 0, U_{tj}^l \geqslant 0 \qquad \forall t \in [T], j \in [m], l \in [k] \tag{B.7e}$$

where $\hat{p}_{tj} = p_{t\sigma(j)}$ and

$$\hat{Q}_j = \min\{\sum_{t=1}^{T} \hat{G}_{tj}, k\}, \forall j < m \text{ and } \hat{Q}_m = \min\{T, k\}$$

Note that $\sigma(m) = m$. We have $\hat{Q}_j = \sum_{t=1}^{j} g_{\sigma(t)}$. We now construct a feasible solution to (B.7) from $\{\Delta_j^*, U_{tj}^{l*}, V_t^{k*}\}$, which is the optimal solution of (B.5) that gives rise to the definition of the permutation $\{\sigma(j), \forall j = 1, \ldots, m\}$. To be specific, we define

$$\hat{V}_t^k = V_t^{k^*}, \hat{U}_{tj}^l = U_{t\sigma(j)}^{l*}, \text{ and } \hat{\Delta}_j = r_{\sigma(j)}^* - r_{\sigma(j+1)}^*, \forall t \in [T], l \in [k], j \in [m]$$

where we define $r_{\sigma(m+1)}^* = 0$. Clearly, constraints (B.5b) and (B.5c) imply that constraints (B.7b) and (B.7c) are satisfied by $\{\hat{\Delta}_j, \hat{U}_{tj}^l, \hat{V}_t^k\}$. Also, we have

$$\sum_{j=1}^{m} \hat{Q}_j \Delta_j = \sum_{j=1}^{m} r_{\sigma(j)}^* \cdot (\hat{Q}_j - \hat{Q}_{j-1}) = \sum_{j=1}^{m} r_{\sigma(j)}^* \cdot g_{\sigma(j)} = 1$$

where the last equation follows from constraint (B.5e). Thus, our proof is completed. □

Proof of Lemma 3.17.1. We prove stronger results by induction. First, we show that for any t, it holds that

$$\sum_{l'=l}^{k} x_t^{l'} = \Pr[\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l+1], \quad \forall l = 1, \ldots, k \tag{B.8}$$

We prove (B.8) by induction on t. When $t = 1$, clearly, for any $l = 1, \ldots, k$, we have

$$\sum_{l'=l}^{k} x_1^{l'} = 1 = \Pr[0 < k-l+1] = \Pr[\sum_{t'<1} \mathrm{Ber}(\tau_{t'}) < k-l+1]$$

which implies that (B.8) holds. We now assume that (B.8) holds for t, and we consider $t+1$. For any $l = 1, \ldots, k$, we have that

$$x_{t+1}^{l} = x_t^{l} - y_t^{l} + y_t^{l+1} = (1-\tau_t) \cdot x_t^{l} + \tau_t \cdot x_t^{l+1}$$

where we denote $x_t^{k+1} = 0$. Thus, we have

$$\sum_{l'=l}^{k} x_{t+1}^{l'} = (1-\tau_t) \cdot \sum_{l'=l}^{k} x_t^{l'} + \tau_t \cdot \sum_{l'=l+1}^{k} x_t^{l'}$$

On the other hand, conditioning whether $\mathrm{Ber}(\tau_t) = 1$, we have

$$\begin{aligned}
\Pr[\sum_{t'<t+1} \mathrm{Ber}(\tau_{t'}) < k-l+1] =& \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l+1) \cdot \Pr[\mathrm{Ber}(\tau_t) = 0] \\
&+ \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l) \cdot \Pr[\mathrm{Ber}(\tau_t) = 1] \\
=& (1-\tau_t) \cdot \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l+1) + \tau_t \cdot \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l)
\end{aligned}$$

From induction hypothesis, we know that

$$\sum_{l'=l}^{k} x_t^{l'} = \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l+1) \text{ and } \sum_{l'=l+1}^{k} x_t^{l'} = \Pr(\sum_{t'<t} \mathrm{Ber}(\tau_{t'}) < k-l)$$

Thus, we have that

$$\sum_{l'=l}^{k} x_{t+1}^{l'} = \Pr[\sum_{t'<t+1} \mathrm{Ber}(\tau_{t'}) < k - l + 1]$$

From induction, we know that (B.8) holds for any $t \in [T]$, which proves the first equation in Lemma 3.17.1.

We now prove the second equation. We prove by induction to show that for any t, it holds

$$\sum_{t'=1}^{t} \tau_{t'} \cdot \sum_{l=1}^{k} x_{t'}^{l} = \mathbb{E}[\min\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}), k\}] \tag{B.9}$$

When $t = 1$, clearly, (B.9) holds. We now assume that (B.9) holds for t and we consider $t + 1$. We have

$$\sum_{t'=1}^{t+1} \tau_{t'} \cdot \sum_{l=1}^{k} x_{t'}^{l} = \sum_{t'=1}^{t} \tau_{t'} \cdot \sum_{l=1}^{k} x_{t'}^{l} + \tau_{t+1} \cdot \sum_{l=1}^{k} x_{t+1}^{l} = \mathbb{E}[\min\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}), k\}] + \tau_{t+1} \cdot \sum_{l=1}^{k} x_{t+1}^{l}$$

On the other hand, denote by $\mathcal{A}_t$ the event $\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}) < k\}$. We have

$$\mathbb{E}[\min\{\sum_{t'=1}^{t+1} \mathrm{Ber}(\tau_{t'}), k\}] = \mathbb{E}[\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}) + \mathrm{Ber}(\tau_{t+1})|\mathcal{A}_t] \cdot \Pr(\mathcal{A}_t) + k \cdot (1 - \Pr(\mathcal{A}_t))$$

Note that the random variable $\mathrm{Ber}(\tau_{t+1})$ is independent of the event $\mathcal{A}_t$. We have

$$\begin{aligned}\mathbb{E}[\min\{\sum_{t'=1}^{t+1} \mathrm{Ber}(\tau_{t'}), k\}] &= \mathbb{E}[\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'})|\mathcal{A}_t] \cdot \Pr(\mathcal{A}_t) + k \cdot (1 - \Pr(\mathcal{A}_t)) + \mathbb{E}[\mathrm{Ber}(\tau_{t+1})] \cdot \Pr(\mathcal{A}_t) \\ &= \mathbb{E}[\min\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}), k\}] + \tau_{t+1} \cdot \Pr(\mathcal{A}_t)\end{aligned}$$

From (B.8), we know that

$$\Pr(\mathcal{A}_t) = \sum_{l=1}^{k} x_{t+1}^{l}.$$

Thus, we have

$$\sum_{t'=1}^{t+1} \tau_{t'} \cdot \sum_{l=1}^{k} x_{t'}^{l} = \mathbb{E}[\min\{\sum_{t'=1}^{t} \mathrm{Ber}(\tau_{t'}), k\}] + \tau_{t+1} \cdot \sum_{l=1}^{k} x_{t+1}^{l} = \mathbb{E}[\min\{\sum_{t'=1}^{t+1} \mathrm{Ber}(\tau_{t'}), k\}]$$

By induction, (B.9) holds for any $t \in [T]$, which completes our proof. □

Proof of Lemma 3.17.2. By (3.13), the objective value of $\mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{ExAnte}}(\mathbf{G})$ for any OST defined by J, ρ will equal

$$\begin{aligned}
\theta &= \min_{j\in[m]} \frac{\sum_{t=1}^{T} \min\{(1-\rho)G_{t,J-1} + \rho G_{tJ}, G_{tj}\} \sum_{l=1}^{k} x_t^l}{\min\{\sum_{t=1}^{n} G_{tj}, k\}} \\
&= \min\left\{ \min_{j<J} \frac{\sum_{t=1}^{n} G_{tj} \sum_{l=1}^{k} x_t^l}{\min\{\sum_{t=1}^{n} G_{tj}, k\}}, \min_{j\geqslant J} \frac{\sum_{t=1}^{T} ((1-\rho)G_{t,J-1} + \rho G_{tJ}) \sum_{l=1}^{k} x_t^l}{\min\{\sum_{t=1}^{n} G_{tj}, k\}} \right\} \\
&\geqslant \min\left\{ \min_{j<J} \frac{\sum_{t=1}^{n} G_{tj} \sum_{l=1}^{k} x_t^l}{\sum_{t=1}^{n} G_{tj}}, \frac{\sum_{t=1}^{T} ((1-\rho)G_{t,J-1} + \rho G_{tJ}) \sum_{l=1}^{k} x_t^l}{k} \right\} \\
&\geqslant \min\left\{ \min_{t\in[T]} \sum_{l=1}^{k} x_t^l, \frac{\sum_{t=1}^{T-1} ((1-\rho)G_{t,J-1} + \rho G_{tJ}) \sum_{l=1}^{k} x_t^l}{k} \right\} \\
&= \min\left\{ \min_{t\in[T]} \Pr\left[\sum_{t'<t} \mathrm{Ber}((1-\rho)G_{t',J-1} + \rho G_{t'J}) < k \right], \frac{\mathbb{E}[\min\{\sum_{t=1}^{T-1} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k} \right\}
\end{aligned}$$

where the first argument in the second inequality holds because $\frac{\sum_{t=1}^{n} G_{tj} \sum_{l=1}^{k} x_t^l}{\sum_{t=1}^{n} G_{tj}} \geqslant \min_{t\in[T]} \sum_{l=1}^{k} x_t^l$ for all $j < J$, and the final equality applies Lemma 3.17.1 throughout the terms. The proof is then completed by the observation that the min over $t \in [T]$ is always achieved when $t = T$. □

Proof of Lemma 3.17.3. Given any distributions $\mathbf{G}$ over m types for the outer problem in (3.14), we construct distributions $\mathbf{G}'_1, \ldots, \mathbf{G}'_T$ over $m + 2$ types such that for some small $\varepsilon > 0$, we have

that $\sup_{J',\rho'} \text{innerLP}_{k,T}^{\text{OST}(J',\rho')/\text{Proph}}(\mathbf{G}'_1,\dots,\mathbf{G}'_T)$ is at most

$$\frac{\varepsilon}{k} + \sup_{J,\rho} \min\left\{ \Pr\left[\sum_{t<T} \text{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T} \text{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k} \right\}. \tag{B.10}$$

Taking $\varepsilon \to 0$ would then complete the proof.

The construction entails defining type distributions for agents $t = 1,\dots,T-1$ as

$$G'_{tj} = \begin{cases} 1, & j = m+2; \\ G_{t,j-1}, & j = 2,\dots,m+1; \\ 0, & j = 1. \end{cases}$$

For the last agent, we define $G'_{nj} = \varepsilon$ for all $j \leqslant m+1$ and $G'_{T,m+2} = 1$. Note that this feasibly satisfies $0 \leqslant G'_{t1} \leqslant \cdots \leqslant G'_{t,m+2} = 1$ for all agents $t \in [T]$.

Recall that for θ to be feasible in $\text{innerLP}_{k,T}^{\text{OST}(J',\rho')/\text{Proph}}(\mathbf{G}'_1,\dots,\mathbf{G}'_T)$, applying (3.13), we need

$$\theta \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G'_{tj}), k\}] \leqslant \sum_{t=1}^{T} \min\{(1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}, G'_{tj}\} \sum_{l=1}^{k} x_t^l \quad \forall j \in [m+2]. \tag{B.11}$$

Taking $j = 1$, we have $\mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G'_{tj}), k\}] = \mathbb{E}[\min\{\text{Ber}(\varepsilon), k\}] = \varepsilon$. Meanwhile, the RHS of (B.11) can be at most $\varepsilon \sum_l x_T^l$ when $j = 1$. Thus, we know that $\theta \leqslant \sum_l x_T^l$. Since the feasible vectors $\mathbf{x}, \mathbf{y}$ in $\text{innerLP}_{k,T}^{\text{OST}(J',\rho')/\text{Proph}}(\mathbf{G}'_1,\dots,\mathbf{G}'_T)$ satisfies the static threshold constraint (3.12) and $(\mathbf{x},\mathbf{y}) \in \mathcal{P}_T^k$, by Lemma 3.17.1, we know $\sum_l x_T^l = \Pr[\sum_{t<T} \text{Ber}((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}) < k]$.

On the other hand, taking $j = m+1$, we have $\mathbb{E}[\min\{\sum_{t=1}^{T} \text{Ber}(G'_{tj}), k\}] = \mathbb{E}[\min\{(T-1) + \text{Ber}(\varepsilon), k\}] = k$ since $G'_{t,m+1} = G_{im} = 1$ for all $t \in [T-1]$ and $T > k$. Meanwhile, the RHS of (B.11)

can be at most $\sum_{t=1}^{T-1}((1-\rho')G'_{t,J'-1}+\rho' G'_{tJ'})\sum_{l=1}^{k} x_t^l + \varepsilon$ when $j = m+1$. Thus, we also know that

$$\begin{aligned}\theta &\leqslant \frac{\varepsilon}{k} + \frac{\sum_{t=1}^{T-1}((1-\rho')G'_{t,J'-1}+\rho' G'_{tJ'})\sum_{l=1}^{k} x_t^l}{k} \\ &= \frac{\varepsilon}{k} + \frac{\mathbb{E}[\min\{\sum_{t=1}^{T-1}\mathrm{Ber}((1-\rho')G'_{t,J'-1}+\rho' G'_{tJ'}), k\}]}{k}\end{aligned}$$

where we have again applied Lemma 3.17.1.

Finally, by setting $J = J' - 1$ and $\rho = \rho'$, due to the construction of $\mathbf{G}'_1, \ldots, \mathbf{G}'_T$ based on $\mathbf{G}$, expression (B.10) evaluates to exactly

$$\frac{\varepsilon}{k} + \min\left\{\Pr\left[\sum_{t<T}\mathrm{Ber}((1-\rho')G'_{t,J'-1}+\rho' G'_{tJ'}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T}\mathrm{Ber}((1-\rho')G'_{t,J'-1}+\rho' G'_{tJ'}), k\}]}{k}\right\}.$$

(Setting $J = J' - 1$ is only valid if $J' \notin \{1, m+2\}$, but it is easy to see that the sup over J', ρ' never requires these values of J' to achieve.) This completes the proof of Lemma 3.17.3. □

Proof of Theorem 3.18. Note that $\mathrm{Proph}(I) \leqslant \mathrm{ExAnte}(I)$ for any instance I, and hence

$$\mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{ExAnte}}(\mathbf{G}) \leqslant \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{Proph}}(\mathbf{G}).$$

Meanwhile, by Lemmas 3.17.2 and 3.17.3, we have

$$\begin{aligned}&\inf_{\mathbf{G}}\sup_{J,\rho}\ \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{ExAnte}}(\mathbf{G}) \\ \geqslant &\inf_{\mathbf{G}}\sup_{J,\rho}\ \min\left\{\Pr\left[\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T}\mathrm{Ber}((1-\rho)G_{t,J-1}+\rho G_{tJ}), k\}]}{k}\right\} \\ \geqslant &\inf_{\mathbf{G}}\sup_{J,\rho}\ \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathrm{Proph}}(\mathbf{G}).\end{aligned}$$

Combining these inequalities completes the proof of Theorem 3.18. □

Proof of Lemma 3.19.1. We construct the instance $\mathbf{G}$ as follows: there are two agents, $T = 2$, four types, $m = 4$, and one slot, $k = 1$. The distribution for the first agent is given by $\mathbf{G}_1 = (0, \frac{1}{2}, \frac{1}{2}, 1)$, and the distribution for the second agent is given by $\mathbf{G}_2 = (\varepsilon, \varepsilon, 1, 1)$.

For any J, ρ, we denote $C_{J,\rho}$ as a vector such that $C_{J,\rho}(j) = \sum_{t=1}^{2} \min\{y_t(J,\rho), G_{tj}x_t(J,\rho)\}$ for each $j \in [m]$, where $\mathbf{x}(J,\rho), \mathbf{y}(J,\rho)$ satisfy

$$x_1(J,\rho) = 1, \qquad y_1(J,\rho) = ((1-\rho)G_{1,J-1} + \rho G_{1J})x_1(J,\rho)$$
$$x_2(J,\rho) = x_1(J,\rho) - y_1(J,\rho), \quad y_2(J,\rho) = ((1-\rho)G_{2,J-1} + \rho G_{2J})x_2(J,\rho).$$

Clearly, we have $C_{1,1} = (\varepsilon, \varepsilon, \varepsilon, \varepsilon)$ and $C_{3,1} = (\frac{\varepsilon}{2}, \frac{1}{2} + \frac{\varepsilon}{2}, 1, 1)$. Moreover, note that

$$Q^{\mathsf{Proph}} = (\varepsilon + o(\varepsilon), \frac{1}{2} + O(\varepsilon), 1, 1) \text{ and } Q^{\mathsf{ExAnte}} = (\varepsilon + o(\varepsilon), \frac{1}{2} + O(\varepsilon), 1, 1).$$

By setting $\mu(1,1) = \frac{1}{3}$, $\mu(3,1) = \frac{2}{3}$, and $\mu(J,\rho) = 0$ for all other J, ρ in $\mathrm{innerLP}_{k,T}^{\mathsf{ST/Proph}}(\mathbf{G})$ and $\mathrm{innerLP}_{k,T}^{\mathsf{ST/ExAnte}}(\mathbf{G})$, we know that

$$\mathrm{innerLP}_{k,T}^{\mathsf{ST/Proph}}(\mathbf{G}) \geq \frac{2}{3} \text{ and } \mathrm{innerLP}_{k,T}^{\mathsf{ST/ExAnte}}(\mathbf{G}) \geq \frac{2}{3}.$$

On the other hand, we show that an OST cannot achieve a guarantee better than $\frac{1}{2}$ with respect to both the Ex-Ante benchmark and the prophet benchmark.

If $J \leq 2$, irregardless of the value of ρ, we have that $y_2(J,\rho) = \varepsilon x_2(J,\rho) \leq \varepsilon$. Then we have that $C_{J,\rho}(3) \leq \min\{y_1(J,\rho), \frac{x_1(J,\rho)}{2}\} + \varepsilon \leq \frac{1}{2} + \varepsilon$. Compared with $Q^{\mathsf{Proph}}(3) = Q^{\mathsf{ExAnte}}(3) = 1$, we know θ cannot be better than $\frac{1}{2}$ as $\varepsilon \to 0$.

If $J \geq 3$, irregardless of the value of ρ, we have that $x_2(J,\rho) = x_1(J,\rho) - y_2(J,\rho) \leq \frac{1}{2}$, and thus $C_{J,\rho}(1) \leq \varepsilon \cdot x_2(J,\rho) \leq \frac{\varepsilon}{2}$. Compared with $Q^{\mathsf{Proph}}(1) = Q^{\mathsf{ExAnte}}(1) = \varepsilon$, we know θ cannot be

better than $\frac{1}{2}$ as $\varepsilon \to 0$. Thus, we have

$$\sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G}) \leq \frac{1}{2} \text{ and } \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G}) \leq \frac{1}{2}$$

as $\varepsilon \to \infty$, which completes our proof. □

Proof of Lemma 3.19.2. The proof mainly follows the proof of Lemma 3.17.3. Given any distributions $\mathbf{G}$ over m types, we construct distributions $\mathbf{G}' = (\mathbf{G}'_1, \ldots, \mathbf{G}'_T)$ over $m+2$ types such that for some small $\varepsilon > 0$, we have that

$$\text{innerLP}_{k,T}^{\text{ST}/\text{Proph}}(\mathbf{G}') \leq \frac{\varepsilon}{k} + \min \left\{ \int_{J,\rho} \Pr\left[\sum_{t<T} \text{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k \right] \mu(J,\rho), \right.$$
$$\left. \int_{J,\rho} \frac{\mathbb{E}[\min\{\sum_{t<T} \text{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k} \mu(J,\rho) \right\}. \tag{B.12}$$

for a probability measure μ over (J,ρ). Taking $\varepsilon \to 0$ would then complete the proof.

The construction entails defining type distributions for agents $t = 1, \ldots, T-1$ as

$$G'_{tj} = \begin{cases} 1, & j = m+2; \\ G_{t,j-1}, & j = 2, \ldots, m+1; \\ 0, & j = 1. \end{cases}$$

For the last agent, we define $G'_{nj} = \varepsilon$ for all $j \leqslant m+1$ and $G'_{T,m+2} = 1$. Note that this feasibly satisfies $0 \leqslant G'_{t1} \leqslant \cdots \leqslant G'_{t,m+2} = 1$ for all agents $t \in [T]$.

Recall that for θ to be feasible in $\text{innerLP}_{k,T}^{\text{ST}/\text{Proph}}(\mathbf{G}'_1, \ldots, \mathbf{G}'_T)$, from constraints (3.8c) and

(3.8b), we have

$$\theta \cdot \mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G'_{tj}), k\}] \leqslant \int_{J',\rho'} \mu'(J',\rho') \cdot \left(\sum_{t=1}^{T} \min\{(1-\rho')G'_{t,J'-1} + \rho G'_{tJ'}, G'_{tj}\} \sum_{l=1}^{k} x_t^l(J',\rho') \right) \quad \forall j \in [m+2] \tag{B.13}$$

for some $(\mathbf{x}(J',\rho'), \mathbf{y}(J',\rho')) \in \mathcal{P}_T^k$ and a measure μ' over $J' \in [m+2]$ and $\rho' \in (0,1]$.

Taking $j = 1$, we have $\mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G'_{t1}), k\}] = \varepsilon$. Meanwhile, when $j = 1$, the RHS of (B.13) can be at most

$$\int_{J',\rho'} \mu'(J',\rho') \cdot \left(\sum_{t=1}^{n} G'_{t1} \sum_{l=1}^{k} x_t^l(J',\rho') \right) = \varepsilon \cdot \int_{J',\rho'} \mu'(J',\rho') \cdot \sum_{l=1}^{k} x_T^l(J',\rho').$$

Thus, we know that

$$\theta \leq \int_{J',\rho'} \mu'(J',\rho') \cdot \sum_{l=1}^{k} x_T^l(J',\rho').$$

Since the feasible vectors $\mathbf{x}(J',\rho'), \mathbf{y}(J',\rho')$ in $\mathrm{innerLP}_{k,T}^{\mathsf{ST/Proph}}(\mathbf{G}'_1, \ldots, \mathbf{G}'_T)$ satisfies the static threshold constraint (3.12) and $(\mathbf{x}(J',\rho'), \mathbf{y}(J',\rho')) \in \mathcal{P}_T^k$, by Lemma 3.17.1, we know $\sum_l x_T^l(J',\rho') = \Pr[\sum_{t<T} \mathrm{Ber}((1-\rho')G'_{t,J-1} + \rho' G'_{tJ}) < k]$ for each J', ρ', which implies that

$$\theta \leq \int_{J',\rho'} \mu'(J',\rho') \cdot \Pr[\sum_{t<T} \mathrm{Ber}((1-\rho')G'_{t,J-1} + \rho' G'_{tJ}) < k]$$

On the other hand, taking $j = m+1$, we have $\mathbb{E}[\min\{\sum_{t=1}^{T} \mathrm{Ber}(G'_{t,m+1}), k\}] = k$ since $G'_{t,m+1} = G_{im} = 1$ for all $t \in [T-1]$ and $T > k$. Meanwhile, the RHS of (B.13) can be at most

$$\int_{J'.\rho'} \mu'(J',\rho') \cdot \left(\sum_{t=1}^{T-1} ((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}) \sum_{l=1}^{k} x_t^l(J',\rho') \right) + \varepsilon$$

when $j = m + 1$. Thus, we also know that

$$\begin{aligned}\theta &\leqslant \frac{\varepsilon}{k} + \frac{\int_{J'.\rho'} \mu'(J',\rho') \cdot \left(\sum_{t=1}^{T-1}((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}) \sum_{l=1}^{k} x_t^l(J',\rho')\right)}{k} \\ &= \frac{\varepsilon}{k} + \int_{J',\rho'} \mu(J',\rho') \cdot \frac{\mathbb{E}[\min\{\sum_{t=1}^{T-1} \mathrm{Ber}((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}), k\}]}{k}\end{aligned}$$

where we have again applied Lemma 3.17.1 to derive the last equality.

Finally, by setting $J = J' - 1$ and $\rho = \rho'$, due to the construction of $\mathbf{G}'_1, \ldots, \mathbf{G}'_T$ based on $\mathbf{G}$, expression (B.12) evaluates to exactly

$$\begin{aligned}\frac{\varepsilon}{k} + \min\Bigg\{ &\int_{J',\rho'} \mu'(J',\rho') \Pr\left[\sum_{t<T} \mathrm{Ber}((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}) < k\right], \\ &\int_{J',\rho'} \mu'(J',\rho') \frac{\mathbb{E}[\min\{\sum_{t<T} \mathrm{Ber}((1-\rho')G'_{t,J'-1} + \rho' G'_{tJ'}), k\}]}{k}\Bigg\}.\end{aligned}$$

(Setting $J = J' - 1$ is only valid if $J' \notin \{1, m+2\}$, but it is easy to see that the sup over J', ρ' never requires these values of J' to achieve.) This completes the proof of Lemma 3.19.2. □

Proof of Theorem 3.20. In Theorem 3.18, we showed that

$$\begin{aligned}&\inf_{\mathbf{G}} \sup_{J,\rho} \ \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathsf{Proph}}(\mathbf{G}) = \inf_{\mathbf{G}} \sup_{J,\rho} \ \mathrm{innerLP}_{k,T}^{\mathrm{OST}(J,\rho)/\mathsf{ExAnte}}(\mathbf{G}) \\ &= \inf_{\mathbf{G}} \sup_{J,\rho} \ \min\left\{\Pr\left[\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}) < k\right], \frac{\mathbb{E}[\min\{\sum_{t<T} \mathrm{Ber}((1-\rho)G_{t,J-1} + \rho G_{tJ}), k\}]}{k}\right\}.\end{aligned}$$

From Lemma 11 in Chawla et al. (2020), the infimum is achieved when there is only $m = 1$ type and $G_{t1} = 1$ for all t, which implies that

$$\begin{aligned}
&\inf_{\mathbf{G}'} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{Proph}}(\mathbf{G}') = \inf_{\mathbf{G}'} \sup_{J,\rho} \text{innerLP}_{k,T}^{\text{OST}(J,\rho)/\text{ExAnte}}(\mathbf{G}') \\
&= \sup_{\rho\in(0,1]} \min \left\{ \Pr\left[\text{Bin}(T-1,\rho) < k\right], \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k} \right\} \\
&= \sup_{\rho\in(0,1]} \min_{\beta\in[0,1]} \beta \cdot \Pr\left[\text{Bin}(T-1,\rho) < k\right] + (1-\beta) \cdot \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k}. \qquad \text{(B.14)}
\end{aligned}$$

Then, fixing $\mathbf{G}$ such that $m = 1$ and $G_{t1} = 1$ for each $t \in [T]$, from Lemma 3.19.2, we have

$$\begin{aligned}
&\inf_{\mathbf{G}'} \text{innerLP}_{k,T}^{\text{ST/ExAnte}}(\mathbf{G}') \le \inf_{\mathbf{G}'} \text{innerLP}_{k,T}^{\text{ST/Proph}}(\mathbf{G}') \\
&\le \sup_{\mu:(0,1]\to\mathbb{R}_{\ge 0}, \int_\rho \mu(\rho)=1} \min \left\{ \int_\rho \Pr\left[\text{Bin}(T-1,\rho) < k\right] \mu(\rho), \int_\rho \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k} \mu(\rho) \right\} \\
&= \min_{\beta\in[0,1]} \sup_{\mu:(0,1]\to\mathbb{R}_{\ge 0}, \int_\rho \mu(\rho)=1} \beta \cdot \int_\rho \Pr\left[\text{Bin}(T-1,\rho) < k\right] \mu(\rho) + (1-\beta) \cdot \int_\rho \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k} \mu(\rho) \\
&= \min_{\beta\in[0,1]} \sup_{\rho\in(0,1]} \beta \cdot \Pr\left[\text{Bin}(T-1,\rho) < k\right] + (1-\beta) \cdot \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k}. \qquad \text{(B.15)}
\end{aligned}$$

Note that OST is a special case of ST, the final result follows as long as the value of (B.14) equals the value of (B.15). Denote by $\lambda(\beta, \rho)$ the function:

$$\lambda(\beta,\rho) := \beta \cdot \Pr\left[\text{Bin}(T-1,\rho) < k\right] + (1-\beta) \cdot \frac{\mathbb{E}[\min\{\text{Bin}(T-1,\rho), k\}]}{k}.$$

It suffices to show that

$$\min_{\beta\in[0,1]} \sup_{\rho\in(0,1]} \lambda(\beta,\rho) = \sup_{\rho\in(0,1]} \min_{\beta\in[0,1]} \lambda(\beta,\rho)$$

Note that $\lambda(\beta, \rho)$ is linear in β, from Sion's minimax theorem, it only remains to show that $\lambda(\beta, \rho)$ is quasi-concave over $\rho \in (0, 1]$, for each fixed $\beta \in [0, 1]$. We now assume β is fixed and we show $\lambda(\beta, \rho)$ is a unimodal function over ρ, which implies quasi-concavity.

Note that

$$\begin{aligned}\lambda(\beta,\rho) &= \beta\sum_{s=0}^{k-1}C_T^s\rho^s(1-\rho)^{T-s}+(1-\beta)\left\{\frac{1}{k}\sum_{s=1}^{k-1}sC_T^s\rho^s(1-\rho)^{T-s}+\sum_{s=k}^{T}C_T^s\rho^s(1-\rho)^{T-s}\right\}\\ &= 1-\beta+(2\beta-1)\sum_{s=0}^{k-1}C_T^s\rho^s(1-\rho)^{T-s}+\frac{1-\beta}{k}\sum_{s=1}^{k-1}sC_T^s\rho^s(1-\rho)^{T-s}\end{aligned}$$

Then, The derivative of $\lambda(\beta,\rho)$ over ρ is

$$\begin{aligned}\frac{\partial}{\partial\rho}\lambda(\beta,\rho) &= (2\beta-1)\left\{-T(1-\rho)^{T-1}+\sum_{s=1}^{k-1}[sC_T^s\rho^{s-1}(1-\rho)^{T-s}-(T-s)C_T^s\rho^s(1-\rho)^{T-s-1}]\right\}\\ &\quad+\frac{1-\beta}{k}\sum_{s=1}^{k-1}\left\{s^2C_T^s\rho^{s-1}(1-\rho)^{T-s}-s(T-s)C_T^s\rho^s(1-\rho)^{T-s-1}\right\}\end{aligned}$$

Notice that

$$\begin{aligned}&\sum_{s=1}^{k-1}[sC_T^s\rho^{s-1}(1-\rho)^{T-s}-(T-s)C_T^s\rho^s(1-\rho)^{T-s-1}]\\ =\ &\sum_{s=0}^{k-2}(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1}-\sum_{s=1}^{k-1}(T-s)C_T^s\rho^s(1-\rho)^{T-s-1}\\ =\ &T(1-\rho)^{T-1}-(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k}+\sum_{s=1}^{k-2}[(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1}-(T-s)C_T^s\rho^s(1-\rho)^{T-s-1}]\\ =\ &T(1-\rho)^{T-1}-(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k}\end{aligned}$$

where the last equality holds since

$$(s+1)C_T^{s+1}=(T-s)C_T^s.$$

Similarly,

$$
\begin{aligned}
& \sum_{s=1}^{k-1}\left\{s^2 C_T^s \rho^{s-1}(1-\rho)^{T-s} - s(T-s)C_T^s \rho^s (1-\rho)^{T-s-1}\right\} \\
= & \sum_{s=0}^{k-2}(s+1)^2 C_T^{s+1}\rho^s(1-\rho)^{T-s-1} - \sum_{s=1}^{k-1} s(T-s)C_T^s\rho^s(1-\rho)^{T-s-1} \\
= & T(1-\rho)^{T-1} - (k-1)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k} + \sum_{s=1}^{k-2}\left((s+1)^2 C_T^{s+1} - s(T-s)C_T^s\right)\rho^s(1-\rho)^{T-s-1} \\
= & T(1-\rho)^{T-1} - (k-1)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k} + \sum_{s=1}^{k-2}(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1} \\
= & \sum_{s=0}^{k-2}(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1} - (k-1)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{\partial}{\partial\rho}\lambda(\beta,\rho) = & -(2\beta-1)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k} \\
& + \frac{1-\beta}{k}\sum_{s=0}^{k-2}(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1} - \frac{1-\beta}{k}(k-1)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k} \\
= & -\left(2\beta-1+\frac{1-\beta}{k}(k-1)\right)(T-k+1)C_T^{k-1}\rho^{k-1}(1-\rho)^{T-k} + \frac{1-\beta}{k}\sum_{s=0}^{k-2}(s+1)C_T^{s+1}\rho^s(1-\rho)^{T-s-1}
\end{aligned}
$$

By dividing both sides by $\rho^{k-1}(1-\rho)^{T-k}$, we know that $\frac{\partial}{\partial\rho}\lambda(\rho) = 0$ is equivalent to

$$
\frac{1-\beta}{k}\sum_{s=0}^{k-2}(s+1)C_T^{s+1}\left(\frac{\rho}{1-\rho}\right)^{s-k+1} = \left(2\beta-1+\frac{1-\beta}{k}(k-1)\right)(T-k+1)C_T^{k-1}. \tag{B.16}
$$

Clearly, when $\beta = 1$, (B.16) does not hold, which implies that $\lambda(\beta,\rho)$ is either non-decreasing, or non-increasing, over ρ. When $\beta \in [0,1)$, the function

$$
\frac{1-\beta}{k}\sum_{s=0}^{k-2}(s+1)C_T^{s+1}t^{s-k+1}
$$

is strictly decreasing in $t \in [0,1]$. Thus, the equation $\frac{\partial}{\partial \rho}\lambda(\rho) = 0$ can have at most one solution is $[0,1]$, which implies that $\lambda(\beta, \rho)$ is unimodal in $[0,1]$. □

Proof of Theorem 3.24. Denote by $\hat{\mathbf{G}}$ the distributions such that the type distribution of each agent t is a uniform distribution over $[0,1]$. Then, from definitions, we have that

$$\text{iidLP}_{k,T}^{\text{DP/Proph}} = \text{innerLP}_{k,T}^{\text{DP/Proph}}(\hat{\mathbf{G}}) \geq \inf_{\mathbf{G}: G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$$

Moreover, denoting by $\{\theta^*, \mathbf{x}^*, \mathbf{y}^*\}$ the optimal solution of $\text{iidLP}_{k,T}^{\text{DP/Proph}}$, it is easy to see that $\{\theta^*, \mathbf{x}^*, \mathbf{y}^*\}$ is a feasible solution to $\text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$ for any $\mathbf{G}$ satisfying $G_{1j} = \cdots = G_{nj}$ for all j. Thus, we know that

$$\text{iidLP}_{k,T}^{\text{DP/Proph}} \leq \inf_{\mathbf{G}: G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$$

which implies that

$$\text{iidLP}_{k,T}^{\text{DP/Proph}} = \inf_{\mathbf{G}: G_{1j}=\cdots=G_{nj}\forall j} \text{innerLP}_{k,T}^{\text{DP/Proph}}(\mathbf{G})$$

Applying the same arguments to $\text{iidLP}_{k,T}^{\text{DP/ExAnte}}$, $\sup_\tau \text{innerLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}}$, $\sup_\tau \text{innerLP}_{k,T}^{\text{OST}(\tau)/\text{ExAnte}}$, $\text{iidLP}_{k,T}^{\text{ST/Proph}}$ and $\text{iidLP}_{k,T}^{\text{ST/ExAnte}}$, we complete our proof. □

Proof of Theorem 3.25. For any $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k$, the objective value of $\text{iidLP}_{k,T}^{\text{DP/ExAnte}}$ will equal

$$\begin{aligned}
\theta &= \inf_{q\in(0,1]} \frac{\sum_{t=1}^T \sum_{l=1}^k \min\{y_t^l, q x_t^l\}}{\min\{nq, k\}} \\
&= \min\left\{ \inf_{q\in(0,k/T]} \frac{\sum_{t=1}^T \sum_{l=1}^k \min\{\frac{y_t^l}{q}, x_t^l\}}{T}, \inf_{q\in(k/T,1]} \frac{\sum_{t=1}^T \sum_{l=1}^k \min\{y_t^l, q x_t^l\}}{k} \right\} \\
&= \frac{\sum_{t=1}^T \sum_{l=1}^k \min\{y_t^l, \frac{k}{T} x_t^l\}}{k}
\end{aligned} \tag{B.17}$$

where we note that $k/T < 1$ since we are assuming $T > k$. Thus, the problem of $\text{iidLP}_{k,T}^{\text{DP/ExAnte}}$ is equivalent to maximizing expression (B.17) subject to $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}_T^k$. From this it is easy to see that the optimal solution involves setting $y_t^l = \frac{k}{T} x_t^l$ for all l and t, equivalent to a static threshold policy with $\tau = k/T$. Therefore, we have

$$\text{iidLP}_{k,T}^{\text{DP/ExAnte}} = \frac{\frac{k}{T} \sum_t \sum_l x_t^l}{k} = \frac{\mathbb{E}[\min\{\text{Bin}(T, k/T), k\}]}{k} \tag{B.18}$$

where the final equality follows from (3.27).

We now show that the same expression results from analyzing $\sup_\tau \text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}}$ or $\sup_\tau \text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{ExAnte}}$. For the first one, by (3.26), the objective value of $\text{iidLP}_{k,T}^{\text{OST}(\tau)/\text{Proph}}$ for any OST τ will equal

$$\begin{aligned}
\theta &= \inf_{q \in (0,1]} \frac{\min\{\tau, q\} \sum_{t=1}^T \sum_{l=1}^k x_t^l}{\mathbb{E}[\min\{\text{Bin}(T, q), k\}]} \\
&= \min\left\{ \inf_{q \in (0,\tau]} \frac{\sum_{t=1}^T \sum_{l=1}^k x_t^l}{\frac{\mathbb{E}[\min\{\text{Bin}(T,q),k\}]}{q}}, \inf_{q \in (\tau,1]} \frac{\tau \sum_{t=1}^T \sum_{l=1}^k x_t^l}{\mathbb{E}[\min\{\text{Bin}(T, q), k\}]} \right\} \\
&= \min\left\{ \frac{\sum_{t=1}^T \sum_{l=1}^k x_t^l}{T}, \frac{\tau \sum_{t=1}^T \sum_{l=1}^k x_t^l}{k} \right\} && \text{(B.19a)} \\
&= \frac{\mathbb{E}[\min\{\text{Bin}(T, \tau), k\}]}{\max\{\tau T, k\}} && \text{(B.19b)}
\end{aligned}$$

where the first argument in (B.19a) results because $\frac{\mathbb{E}[\min\{\text{Bin}(T,q),k\}]}{q}$ is maximized over $q \in (0, \tau]$ as q approaches 0 from the positive side, and the final equality applies (3.27) to both arguments after multiplying and dividing the first argument by τ. From this it is easy to see that the supremum of expression (B.19b) over τ is achieved by setting $\tau = k/T$, since ratio $\frac{\mathbb{E}[\min\{\text{Bin}(T,\tau),k\}]}{\tau}$ is decreasing over $\tau \in [k/T, 1]$. To see this, by noting that $\mathbb{E}[\min\{\text{Bin}(T, 0), k\}] = 0$, it is enough to show that $\mathbb{E}[\min\{\text{Bin}(T, \tau), k\}$ is concave over $\tau \in [0, 1]$. We define the function $h(x) = \min\{x, k\}$. Then,

we have

$$\begin{aligned}\frac{\partial}{\partial\tau}\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\} &= \sum_{j=0}^{k-1}(k-j)\cdot C_T^j\cdot(j\tau^{j-1}(1-\tau)^{T-j}-(T-j)\tau^j(1-\tau)^{T-j-1})\\ &= T\cdot\sum_{j=0}^{k-2}(k-1-j)\cdot C_{T-1}^j\cdot\tau^j(1-\tau)^{T-1-j}-T\cdot\sum_{j=0}^{k-1}(k-j)\cdot C_{T-1}^j\cdot\tau^j(1-\tau)^{T-1-j}\\ &= T\cdot\mathbb{E}_{X\sim\mathrm{Bin}(T-1,\tau)}[h(X+1)-h(X)]\end{aligned}$$

which implies that

$$\frac{\partial^2}{\partial\tau^2}\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}=T(T-1)\cdot\mathbb{E}_{X\sim\mathrm{Bin}(T-2,\tau)}[h(X+2)+h(X)-2h(X+1)]\le 0$$

by noting that $h(X+2)+h(X)-2h(X+1)\le 0$ for any X. Thus, $\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}$ is concave over $\tau\in[0,1]$ and $\frac{\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}]}{\tau}$ is decreasing over $\tau\in[k/T,1]$.

Therefore, we have shown that $\sup_\tau \mathrm{iidLP}_{k,T}^{\mathrm{OST}(\tau)/\mathrm{Proph}}$ is identical to (B.18).

Finally, the same argument can be made to show that $\sup_\tau \mathrm{iidLP}_{k,T}^{\mathrm{OST}(\tau)/\mathrm{ExAnte}}$ is equal to (B.18), since the same expression (B.19b) can be derived for $\mathrm{iidLP}_{k,T}^{\mathrm{OST}(\tau)/\mathrm{ExAnte}}$. This completes the proof of Theorem 3.25. □

Proof of Theorem 3.26. Let τ be the unique value in $(\underline{\tau},\overline{\tau})$ at which

$$\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}]=\frac{\mathbb{E}[\min\{\mathrm{Bin}(T,\underline{\tau}),k\}]+\mathbb{E}[\min\{\mathrm{Bin}(T,\overline{\tau}),k\}]}{2}. \tag{B.20}$$

We show that for this value of τ, (3.28) holds for all choices of q.

First, if $q>\overline{\tau}$, then all agents accepted by any static threshold τ, $\underline{\tau}$, or $\overline{\tau}$ will have type q or better. In this case, using the fact that $\min\{y_t^l(\tau),qx_t^l(\tau)\}=y_t^l(\tau)=\tau x_t^l(\tau)$ (due to (3.25c), and an analogous argument can be made for $\underline{\tau},\overline{\tau}$) and applying the second identity in (3.27) with $t=T$, we see that (3.28) is equivalent to $\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}]\geqslant\frac{1}{2}(\mathbb{E}[\min\{\mathrm{Bin}(T,\underline{\tau}),k\}]+$

$\mathbb{E}[\min\{\text{Bin}(T,\overline{\tau}),k\}]$). In other words, we need to check that τ accepts no fewer agents than the average of $\underline{\tau}$ and $\overline{\tau}$, which in fact holds as equality due to (B.20).

Next, if $q \in (\tau, \overline{\tau}]$, then all agents accepted by static thresholds $\tau, \underline{\tau}$ still have type q or better, while some of the agents accepted by $\overline{\tau}$ may not. Since τ accepts no fewer agents than the average of $\underline{\tau}$ and $\overline{\tau}$, (3.28) must still be true.

The third case we consider is $q < \underline{\tau}$. Agents of types q or better are accepted by all policies, so the probability of accepting any agent t who has type q or better is simply q times the probability of having a slot available at that time. Formally, using the fact that $\min\{y_t^l(\tau), qx_t^l(\tau)\} = qx_t^l(\tau)$ (due to (3.25c), and an analogous argument can be made for $\underline{\tau}, \overline{\tau}$) and applying the first identity in (3.27), we see that (3.28) is equivalent to

$$q\sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\tau)<k] \geqslant \frac{q}{2}\left(\sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\underline{\tau})<k] + \sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\overline{\tau})<k]\right). \quad \text{(B.21)}$$

In words, after canceling out the q's, we need to prove that the total probability of having a slot available is higher for τ, as compared to the average of $\underline{\tau}$ and $\overline{\tau}$.

Before proving (B.21), we show that the final case $q \in (\underline{\tau}, \tau]$ reduces to it. Indeed, in the final case static threshold $\underline{\tau}$ may not accept some agents who have type q or better, which only decreases the RHS of (B.21). Therefore, establishing (B.21) would complete the proof.

To establish (B.21), we observe that (B.20), which holds by construction, is equivalent to

$$\tau\sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\tau)<k] = \frac{1}{2}\left(\underline{\tau}\sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\underline{\tau})<k] + \overline{\tau}\sum_{t=1}^{T}\Pr[\text{Bin}(t-1,\overline{\tau})<k]\right). \quad \text{(B.22)}$$

Recalling that $\underline{\tau} < \overline{\tau}$, we now argue two facts. First, we immediately see that $\Pr[\text{Bin}(t-1,\underline{\tau}) < k] \geqslant \Pr[\text{Bin}(t-1,\overline{\tau}) < k]$ for all t. Second, we argue that $\tau \leqslant (\underline{\tau}+\overline{\tau})/2$. This is because the function $\mathbb{E}[\min\{\text{Bin}(T,\tau),k\}]$ is strictly increasing and concave in τ—so if $\tau > (\underline{\tau}+\overline{\tau})/2$, then Jensen's

inequality would say $\mathbb{E}[\min\{\mathrm{Bin}(T,\tau),k\}] > \mathbb{E}[\min\{\mathrm{Bin}(T,\frac{\tau+\overline{\tau}}{2}),k\}] \geqslant \frac{1}{2}(\mathbb{E}[\min\{\mathrm{Bin}(T,\underline{\tau}),k\}] + \mathbb{E}[\min\{\mathrm{Bin}(T,\overline{\tau}),k\}])$, which contradicts (B.20). From these facts we derive

$$\begin{aligned}\frac{\underline{\tau}+\overline{\tau}}{2}\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\tau)<k] &\geqslant \tau\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\tau)<k]\\ &= \frac{1}{4}\left(2\underline{\tau}\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\underline{\tau})<k] + 2\overline{\tau}\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\overline{\tau})<k]\right)\\ &\geqslant \frac{1}{4}(\underline{\tau}+\overline{\tau})\left(\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\underline{\tau})<k] + \sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\overline{\tau})<k]\right)\end{aligned}$$

where the equality holds by (B.22), and the inequality holds by applying the rearrangement inequality with $\underline{\tau} < \overline{\tau}$, $\sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\underline{\tau})<k] \geqslant \sum_{t=1}^{T}\Pr[\mathrm{Bin}(t-1,\overline{\tau})<k]$ and then factoring. This implies (B.21) and completes the proof of Theorem 3.26. □

Proof of Lemma 3.27.1. We consider further relaxing LP (3.33). It is clear to see that the expression in large parentheses at the LHS of (3.33b),

$$(1-(1-\kappa)^{T})\theta - \kappa\sum_{t=I}^{T-1}(1-Y_t),$$

is a concave function over κ, which implies that the maximum is attained by setting the first derivative to 0. Letting

$$x_I := \sum_{t=I}^{T-1}(1-Y_t) \qquad z_I := \frac{x_I}{T\theta} \qquad\qquad \forall I = 0,\ldots,T-1$$

we have that the maximum for constraint (3.33b) for $I = 0,\ldots,T-1$ is attained when

$$(1-\kappa)^{T-1} = z_I \qquad \kappa = 1 - z_I^{1/(T-1)}.$$

Note that $Y_I = 1 + x_{I+1} - x_I = 1 + T\theta(z_{I+1} - z_I)$ for all $I = 0, \ldots, T-1$, with $z_T = 0$. Then, constraint (3.33c) can be transformed into the following constraint over the z_I variables for $I = 0, \ldots, T-1$:

$$z_1 - z_0 \leqslant -\frac{1}{T\theta} \leqslant z_2 - z_1 \leqslant \cdots \leqslant z_T - z_{T-1} \leqslant 0.$$

Moreover, the constraints (3.33b) are satisfied if for all $I = 0, \ldots, T-1$,

$$(1 - z_I^{T/(T-1)})\theta - \left(1 - z_I^{1/(T-1)}\right) x_I \leq 1 + x_{I+1} - x_I,$$

which is equivalent to

$$1 - \frac{1}{\theta} + (T-1) z_I^{T/(T-1)} \leq T z_{I+1}.$$

As a result, LP (3.33) can again be characterized as the following optimization problem over z-variables:

$$\begin{aligned}
&\max \theta \\
&\text{s.t.} \quad (T-1) z_I^{T/(T-1)} \leqslant n z_{I+1} + \frac{1}{\theta} - 1 && \forall I = 0, \ldots, T-1 && \text{(B.23a)} \\
&\qquad z_T = 0 \\
&\qquad z_1 - z_0 \leqslant -\frac{1}{T\theta} \leqslant z_2 - z_1 \leqslant \cdots \leqslant z_T - z_{T-1} \leqslant 0 && && \text{(B.23b)}
\end{aligned}$$

Finally, combining the constraint $z_1 - z_0 \leqslant -\frac{1}{T\theta}$ from (B.23b) and the constraint (B.23a) with $I = 0$, we get

$$(T-1) z_0^{T/(T-1)} \leq n z_1 + \frac{1}{\theta} - 1 \leq n z_0 - 1 \qquad \Longrightarrow \qquad (T-1) z_0^{T/(T-1)} - n z_0 \leq -1.$$

Note that expression $(T-1) z_0^{T/(T-1)} - n z_0$ as a function over $z_0 \geqslant 0$ is always at least -1, with equality achieved only when $z_0 = 1$. Therefore, we can add the constraint $z_0 = 1$ to $\mathsf{LP}_T^{\mathsf{mon}}$

without changing the objective value. Together with $z_0 = 1$ and $z_T = 0$, the constraint (B.23b) can be further relaxed into $z_t \in [0, 1]$ for $t = 1, \ldots, T-1$, which establishes $\mathsf{LP}_T^{\text{relax}}$ as a relaxation of LP (B.23). Thus, our proof is completed. □

Proof of Lemma 3.27.2. Denote by $\{\theta, z_I\}_{I=0}^T$ as an optimal solution of $\mathsf{LP}_T^{\text{relax}}$ and denote by I_1 the smallest index such that

$$z_{I_1+1} > \max\{\frac{T-1}{T} z_{I_1}^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T}, 0\}$$

Then, we construct another solution $\{\hat{\theta}, \hat{z}_I\}_{I=0}^T$ such that

$$\hat{\theta} = \theta, \;\; \hat{z}_{I_1+1} = \max\{\frac{T-1}{T} z_{I_1}^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T}, 0\}, \;\; \hat{z}_I = z_I \text{ for } I = 0, \ldots, I_1, I_1+2, \ldots, T$$

Clearly, since $\hat{z}_{I_1+1} \le z_{I_1+1}$, we must have $\{\hat{\theta}, \hat{z}_I\}_{I=0}^T$ as an optimal solution to $\mathsf{LP}_T^{\text{relax}}$. Every time we repeat the above construction procedure, the value of I_1 will be increased by at least 1. Thus, after a finite number of steps, we obtain an optimal solution $\{\hat{\theta}, \hat{z}_I\}_{I=0}^T$ such that

$$\hat{z}_{I+1} = \max\{\frac{T-1}{T} \hat{z}_I^{T/(T-1)} - \frac{1}{T\hat{\theta}} + \frac{1}{T}, 0\}, \;\; \forall I = 0, \ldots, T-1 \tag{B.24}$$

We regard $\hat{z}_I(\theta)$ as a function of θ, for each $I = 0, \ldots, T-1$, where $\hat{z}_I(\theta)$ is computed iteratively from (B.24). Note that the RHS of (B.24) is non-decreasing over θ. We must have $\hat{z}_I(\theta)$ is a non-decreasing function over θ. If there exists an $I_2 \le T-1$ such that $\frac{T-1}{T} \hat{z}_{I_2}(\theta)^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T} < 0$, we can always increase the value of θ by $\varepsilon > 0$ such that it still holds $\frac{T-1}{T} \hat{z}_{I_2}(\theta+\varepsilon)^{T/(T-1)} - \frac{1}{T(\theta+\varepsilon)} + \frac{1}{T} < 0$, and $\{\theta + \varepsilon, \hat{z}_I(\theta+\varepsilon)\}$ is still feasible to $\mathsf{LP}_T^{\text{relax}}$. Thus, in order for $\hat{\theta}$ to be optimal, we must have $\hat{z}_{I+1} = \frac{T-1}{T} \hat{z}_I^{T/(T-1)} - \frac{1}{T\hat{\theta}} + \frac{1}{T}$ for all $I = 0, \ldots, T-1$, which completes our proof. □

Proof of Lemma 3.27.3. Denote $\{\theta, z_I\}_{I=0}^{T}$ as the solution constructed in Lemma 3.27.2. Then, we denote

$$Y_I = 1 + T\theta(z_{I+1} - z_I) \text{ for } I = 0, 1, \ldots, T-1.$$

Clearly, it holds that

$$\max_{\kappa\in(0,1]} \left((1-(1-\kappa)^T)\theta - \kappa \sum_{t=I}^{T-1} (1 - Y_t) - Y_I \right) = 0 \text{ for } I = 0, 1, \ldots, T-1.$$

We also denote $y_t = Y_t - Y_{t-1}$ for $t = 1, \ldots, T-1$. Note that $\sum_{t=1}^{T-1} y_t = Y_{T-1} \le 1$, we also denote $y_T = 1 - \sum_{t=1}^{T-1} y_t$. Then, we have

$$\max_{\kappa\in(0,1]} \left((1-(1-\kappa)^T)\theta - \kappa \sum_{t=I+1}^{T} (1 - \sum_{t'=1}^{t-1} y_{t'}) - \sum_{t=1}^{I} y_t \right) = 0 \text{ for } I = 0, 1, \ldots, T-1. \tag{B.25}$$

In order to show that $\{\theta, y_t\}_{t=1}^{T}$ is a feasible solution to $\text{iidLP}_{1,T}^{\text{DP/Proph}}$, it suffices to show that

$$(1-(1-\kappa)^T)\theta \le \sum_{t\in S} y_t + \kappa \cdot \sum_{t\in[T]\setminus S} (1 - \sum_{t'=1}^{t-1} y_{t'}), \text{ for any } \kappa \in [0,1] \text{ and } S \subset [T] \tag{B.26}$$

where the constraint $y_t \ge 0$ also follows from (B.26) by setting $\mathcal{S} = \{j\}$ and $\kappa = 0$.

We now proceed to prove (B.26). For any fixed S, clearly, the left hand side of constraint (B.26) is a concave function over κ. Thus, after we maximize over κ in (B.26), we get that $\{\theta, y_t\}_{t=1}^{T}$ is a feasible solution to $\text{iidLP}_{1,T}^{\text{DP/Proph}}$ if

$$f(S) \le 0, \quad \forall S \subset [T] \tag{B.27}$$

where $f(S)$ is a set function defined for any subset $S \subset [T]$ as follows

$$f(S) := 1 + (T-1) \cdot \left(\frac{\sum_{t\in[T]\setminus S} \alpha_t}{T\theta} \right)^{\frac{T}{T-1}} - \frac{\sum_{t\in[T]\setminus S} \alpha_t}{\theta} - \frac{\sum_{t\in S} y_t}{\theta}, \text{ for any } S \subset [T] \tag{B.28}$$

and we denote $1 - \sum_{t'=1}^{t-1} y_{t'}$ by α_t for notation brevity. Note that from (B.25) by setting $\kappa = 0$, we have $\sum_{t=1}^{I} y_t \geq 0$ for each $I = 0, 1, \ldots, T-1$. Also, we have

$$\sum_{t=1}^{I} y_t = Y_I - Y_0 = 1 + T\theta(z_{I+1} - z_I) \leq 1, \ \forall I = 0, 1, \ldots, T-1$$

by noting that z_I is a decreasing sequence in I. Thus, we claim that $\alpha_t \in [0, 1]$ for each $t \in [T]$.

The condition (B.25) can also be expressed via the function $f(\cdot)$. We denote by $E_t = \{1, 2, \ldots, t\}$ for each $t \in [T]$. Note that the left hand side of (B.25) is a concave function over κ. Then after maximizing over κ in (B.25), we have that

$$f(E_t) = 0 \text{ for } t = 1, \ldots, T-1 \text{ and } f(E_T) \leq 0 \tag{B.29}$$

We now use the condition (B.29) to prove (B.27). A key step is to show the set function $f(\cdot)$ to be supermodular. This allows us to ultimately show that it is maximized when S takes the form of an interval $\{1, \ldots, I\}$, for which we already knew by the construction in (B.29).

Note that we have

$$f(S) = g(\sum_{t \in S} \alpha_t) - \frac{\sum_{t \in [T] \setminus S} \alpha_t}{\theta} - \frac{\sum_{t \in S} y_t}{\theta}$$

where

$$g(x) = 1 + (T-1) \cdot \left(\frac{\sum_{t \in [T]} \alpha_t - x}{T\theta} \right)^{\frac{T}{T-1}}$$

It is clear to see that $g(x)$ is a convex function over x. Then, it is well-known (e.g. Lemma 2.6.2 in Topkis (2011)) that $f(S)$ is a supermodular function.

For any set $S \subset \mathcal{T}$, we assume without loss of generality that the elements in S are sorted in an increasing order. We denote $S(t)$ as the t-th element of S and denote by $\sigma(S)$ the number of t such that $S(t+1) > S(t) + 1$. Then, for any k, we denote

$$\mathcal{T}_k = \{S : \sigma(S) \leq k\}$$

Clearly, we have $T_0 = \{E_1, \ldots, E_T\}$, which implies that $\max_{S \in T_0} f(S) \leq 0$. Now we will prove (B.27) by induction. Suppose that there exists an integer l such that

$$\max_{S \in T_l} f(S) \leq 0$$

For any set $\hat{S} \subset [T]$ such that $\sigma(\hat{S}) = l + 1$, we denote $\hat{t}$ as the largest index such that $\hat{S}(\hat{t} + 1) > \hat{S}(\hat{t}) + 1$, and we denote $\hat{S}(\hat{j})$ as the last element of the set $\hat{S}$. We denote $\hat{T} = E_{\hat{S}(t)}$. Clearly, it holds that

$$\hat{T} \cup \hat{S} = E_{\hat{S}(\hat{j})} \text{ and } \sigma(\hat{T} \cap \hat{S}) = l$$

From the supermodularity of $f(S)$, we have

$$f(\hat{T}) + f(\hat{S}) \leq f(\hat{T} \cup \hat{S}) + f(\hat{T} \cap \hat{S})$$

From the induction hypothesis, we know $f(\hat{T} \cap \hat{S}) \leq 0$. Also, from (B.29), we know $f(\hat{T} \cup \hat{S}) = f(E_{\hat{S}(\hat{j})}) \leq 0$ and $f(\hat{T}) = f(E_{\hat{S}(\hat{t})}) = 0$ since $\hat{S}(\hat{t}) \leq T - 1$. Thus, we concludes that $f(\hat{S}) = 0$, which implies that

$$\max_{S \subset T_{l+1}} f(S) \leq 0$$

From the induction, we know that

$$\max_{S \subset T_T} f(S) = \max_{S \subset [T]} f(S) \leq 0$$

which completes our proof. □

Proof of Lemma 3.27.4. We denote $\{\theta, z_I\}_{I=0}^{2T}$ as an optimal solution to $\mathsf{LP}_{2T}^{\text{relax}}$. We construct a feasible solution to $\mathsf{LP}_{T}^{\text{relax}}$, denoted by $\{\hat{\theta}, \hat{z}_I\}_{I=0}^{T}$, To be specific, we set

$$\hat{\theta} = \theta, \quad \hat{z}_I = z_{2I} \text{ for } I = 0, 1, \ldots, T$$

It only remains to show that

$$\hat{z}_{I+1} = z_{2I+2} \geq \frac{T-1}{T} z_{2I}^{T/(T-1)} - \frac{1}{T\theta} + \frac{1}{T}, \quad \forall I = 0, \ldots, T-1$$

Note that we have

$$z_{2I+2} \geq \frac{(2T-1)\left[\frac{(2T-1)z_{2I}^{\frac{2T}{2T-1}} - \frac{1}{\theta} + 1}{2T}\right]^{\frac{2T}{2T-1}} - \frac{1}{\theta} + 1}{2T}$$

Denote by $\alpha = \frac{1}{\theta} - 1$. It only remains to show that

$$f(\alpha) = (2T-1)\left[\frac{(2T-1)z_{2I}^{\frac{2T}{2T-1}} - \alpha}{2T}\right]^{\frac{2T}{2T-1}} - 2(T-1)z_{2I}^{\frac{T}{T-1}} + \alpha \geq 0$$

Note that

$$f'(\alpha) = 1 - \left[\frac{(2T-1)z_{2I}^{\frac{2T}{2T-1}} - \alpha}{2T}\right]^{\frac{1}{2T-1}}$$

Further note that

$$1 \geq z_{2I+1} \geq \frac{(2T-1)z_{2I}^{\frac{2T}{2T-1}} - \alpha}{2T}$$

It holds that $f'(\alpha) \geq 0$. Thus, in order to show that $f(\alpha) \geq 0$, it suffices to show that $f(0) \geq 0$, t.e.,

$$\left[\frac{(2T-1)z_{2I}^{\frac{2T}{2T-1}}}{2T}\right]^{\frac{2T}{2T-1}} \geq \frac{2T-2}{2T-1} \cdot z_{2I}^{\frac{T}{T-1}}$$

which is equivalent to showing

$$\left(\frac{2T-1}{2T}\right)^{\frac{2T}{2T-1}} \geq \frac{2T-2}{2T-1} \cdot z_{2I}^{\frac{T}{(T-1)(2T-1)^2}}$$

Thus, it only remains to show that

$$(1-\frac{1}{2T})^{2T} \geq (1-\frac{1}{2T-1})^{2T-1} \cdot z_{2I}^{\frac{T}{(T-1)(2T-1)}}$$

The above inequality holds by noting that $z_{2I} \in [0,1]$ and $(1-\frac{1}{2T})^{2T} \geq (1-\frac{1}{2T-1})^{2T-1}$, which completes our proof. □

Proof of Theorem 3.28. We denote function

$$f(x) = x(\ln x - 1) \text{ and } f_T(x) = (T-1) \cdot x^{T/(T-1)} - nx \text{ for each } T$$

Then, for each T, we define a sequence of values $\{H_{T,t}\}_{t=0}^{T}$ such that

$$H_{T,0} = 1,\ H_{T,t} = H_{T,t-1} + \frac{1}{T} \cdot (f_T(H_{T,t-1}) - \frac{1}{\theta_T} + 1) \text{ for } t = 1, \ldots, T$$

where θ_T is selected such that $H_{T,T} = 0$. Clearly, from Lemma 3.27.2, we know that $\{\theta_T, H_{T,t}\}_{t=0}^{T}$ is an optimal solution to $\mathsf{LP}_T^{\mathsf{relax}}$. Then, from Lemma 3.27.3, we know that $\theta_T = \mathrm{iidLP}_{1,T}^{\mathsf{DP/Proph}}$. Moreover, Lemma 3.27.4 implies that $\inf_T \theta_T = \liminf_{T\to\infty} \theta_T$. Thus, it only remains to show that

$$\lim_{T\to\infty} \theta_T = \theta^*$$

The remaining part mainly follows the proof of Lemma 6.2 in **?**. Here, we include the whole proof for completeness. For any $t = 1, \ldots, T$, it holds that

$$H_{T,t} - H_{T,t-1} = \frac{1}{T} \cdot (f(H_{T,t-1}) - \frac{1}{\theta^*} + 1) + \frac{1}{T} \cdot (f_T(H_{T,t-1}) - f(H_{T,t-1})) + \frac{1}{T} \cdot (\frac{1}{\theta^*} - \frac{1}{\theta_T})$$

which implies that

$$\frac{1}{T}\cdot\frac{\frac{1}{\theta^*}-\frac{1}{\theta_T}}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}=\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}-\frac{1}{T}-\frac{1}{T}\cdot\frac{f_T(H_{T,t-1})-f(H_{T,t-1})}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}$$

Note that $f(H_{T,t-1})-\frac{1}{\theta^*}+1\in[-\frac{1}{\theta^*},1-\frac{1}{\theta^*}]$.

Case I. If $\theta^*\le\theta_T$, we have

$$\begin{aligned}\frac{\theta^*}{T}\cdot|\frac{1}{\theta^*}-\frac{1}{\theta_T}|&\le-\frac{1}{T}\cdot\frac{\frac{1}{\theta^*}-\frac{1}{\theta_T}}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}=-\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}+\frac{1}{T}+\frac{1}{T}\cdot\frac{f_T(H_{T,t-1})-f(H_{T,t-1})}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}\\&\le-\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}+\frac{1}{T}+\left|\frac{1}{T}\cdot\frac{f_T(H_{T,t-1})-f(H_{T,t-1})}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}\right|\\&\le-\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}+\frac{1}{T}+\frac{\frac{1}{\theta^*}-1}{T}\cdot\|f_T-f\|_\infty\end{aligned}\tag{B.30}$$

Sum over both sides of (B.30) for $t=1,\dots,T$, we have

$$\theta^*\cdot|\frac{1}{\theta^*}-\frac{1}{\theta_T}|\le-\sum_{t=1}^{T}\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}+1+(\frac{1}{\theta^*}-1)\cdot\|f_T-f\|_\infty$$

Case II. If $\theta^*>\theta_T$, we have

$$\begin{aligned}\frac{\theta^*}{T}\cdot|\frac{1}{\theta^*}-\frac{1}{\theta_T}|&\le\frac{1}{T}\cdot\frac{\frac{1}{\theta^*}-\frac{1}{\theta_T}}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}=\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}-\frac{1}{T}-\frac{1}{T}\cdot\frac{f_T(H_{T,t-1})-f(H_{T,t-1})}{f(H_{T,t-1})-\frac{1}{\theta^*}+1}\\&\le\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}-\frac{1}{T}+\frac{1}{T\theta^*}\cdot\|f_T-f\|_\infty\end{aligned}$$

Sum over both sides for $t=1,\dots,T$, we have

$$\theta^*\cdot|\frac{1}{\theta^*}-\frac{1}{\theta_T}|\le\sum_{t=1}^{T}\frac{H_{T,t-1}-H_{T,t}}{\frac{1}{\theta^*}-1-f(H_{T,t-1})}-1+\frac{1}{\theta^*}\cdot\|f_T-f\|_\infty$$

Thus, on both cases, we have that

$$\theta^* \cdot |\frac{1}{\theta^*} - \frac{1}{\theta_T}| \leq \left| 1 - \sum_{t=1}^{T} \frac{H_{T,t-1} - H_{T,t}}{\frac{1}{\theta^*} - 1 - f(H_{T,t-1})} \right| + \frac{1}{\theta^*} \cdot \|f_T - f\|_\infty$$

Further note that we have $|H_{T,t-1} - H_{T,t}| \leq \frac{1}{T\theta_T}$, which implies

$$\lim_{T\to\infty} \sum_{t=1}^{T} \frac{H_{T,t-1} - H_{T,t}}{\frac{1}{\theta^*} - 1 - f(H_{T,t-1})} = \int_{h=0}^{1} \frac{dh}{\frac{1}{\theta^*} - 1 - f(h)} = 1.$$

Also, note that $\|f_T - f\|_\infty \leq \frac{1}{en}$, then we must have

$$\lim_{T\to\infty} \theta_T = \theta^*$$

which completes our proof. □

C | Appendix for Chapter 4

Proofs of Section 4.2

Proof of Lemma 4.1.1:. Given the realized parameters $\mathcal{H} = (\theta_1, \theta_2, \ldots, \theta_T)$, we can denote the offline optimum of (PCP) as a function of $\mathcal{H}$, namely, $\{x_t^*(\mathcal{H})\}_{t=1}^T$. Let

$$\tilde{x}_t(\theta) = \mathbb{E}\left[x_t^*(\mathcal{H}) | \theta_t = \theta\right]$$

where the conditional expectation is taken with respect to $\mathcal{P}_j$ for $j \neq t$. We show that $\tilde{x}_t(\theta)$ is a feasible solution to (4.4). Specifically, note that for each $i = 1, ..., m$,

$$\begin{aligned} c_i &\geq \mathbb{E}\left[\sum_{t=1}^T g_i(x_t^*(\mathcal{H}); \theta_t)\right] = \sum_{t=1}^T \mathbb{E}_{\theta_t \sim \mathcal{P}_t}\left[\mathbb{E}[g_i(x_t^*(\mathcal{H}); \theta_t) | \theta_t = \theta]\right] \\ &\geq \sum_{t=1}^T \int_{\theta \in \Theta} g_i(\tilde{x}_t(\theta); \theta) d\mathcal{P}_t(\theta) = \sum_{t=1}^T \mathcal{P}_t g_i(x_t; \theta) \end{aligned}$$

where the first inequality comes from the feasibility of the optimal solution $x_t^*(\mathcal{H})$ and the second inequality follows from that the function $g_i(\cdot; \theta_t)$ is a convex function for each i and $\theta_t \in \Theta$. Thus,

$\{\tilde{\boldsymbol{x}}_t(\boldsymbol{\theta})\}$ is a feasible solution to (4.4). Similarly, we can analyze the objective function

$$\begin{aligned}\mathbb{E}[R_T^*] &= \mathbb{E}\left[\sum_{t=1}^{T} f(\boldsymbol{x}_t^*(\mathcal{H}); \boldsymbol{\theta}_t)\right] = \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}_t \sim \mathcal{P}_t}\left[\mathbb{E}[f(\boldsymbol{x}_t^*(\mathcal{H}); \boldsymbol{\theta}_t) | \boldsymbol{\theta}_t = \boldsymbol{\theta}]\right] \\ &\leq \sum_{t=1}^{T} \int_{\boldsymbol{\theta} \in \Theta} f(\tilde{\boldsymbol{x}}_t(\boldsymbol{\theta}); \boldsymbol{\theta}) d\mathcal{P}_t(\boldsymbol{\theta}) \leq R_T^{\mathrm{UB}}\end{aligned}$$

where the first inequality follows from that the function $f(\cdot; \boldsymbol{\theta})$ is a concave function for any $\boldsymbol{\theta} \in \Theta$ and the last inequality comes from the optimality of R_T^{UB}. Thus we complete the proof. □

Proof of Proposition 4.2:. We first prove that $\boldsymbol{p}^*$ is an optimal solution for L_t. Note that for each t, $L_t(\boldsymbol{p})$ is a convex function over $\boldsymbol{p}$ and

$$\nabla L_t(\boldsymbol{p}^*) = \boldsymbol{\gamma}_t + \mathcal{P}_t \nabla h(\boldsymbol{p}^*; \boldsymbol{\theta}_t) = \boldsymbol{\gamma}_t - \mathcal{P}_t \boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t)$$

where $\boldsymbol{x}^*(\boldsymbol{\theta}) = \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}}\{f(\boldsymbol{x}; \boldsymbol{\theta}) - (\boldsymbol{p}^*)^T \cdot \boldsymbol{g}(\boldsymbol{x}; \boldsymbol{\theta})\}$. With the definition of $\boldsymbol{\gamma}_t$ in (4.6), it follows immediately that

$$\nabla L_t(\boldsymbol{p}^*) = 0$$

which implies that $\boldsymbol{p}^*$ is a minimizer of the function $L(\cdot)$ for each t. We then prove that $L(\boldsymbol{p}^*) = \sum_{t=1}^{T} L_t(\boldsymbol{p}^*)$. Define the set of binding constraints $\mathcal{I}_B = \{i : p_i^* > 0, i = 1, ..., m\}$. From the convexity of the function $L(\boldsymbol{p})$ over $\boldsymbol{p}$, for each $i \in \mathcal{I}_B$, it holds that

$$0 = \nabla_i L(\boldsymbol{p}^*) = c_i - \sum_{t=1}^{T} \mathcal{P}_t \boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\theta}_t); \boldsymbol{\theta}_t) = c_i - \sum_{t=1}^{T} \gamma_{t,i}$$

Thus, we have that

$$\boldsymbol{c}^\top \cdot \boldsymbol{p}^* = \sum_{t=1}^{T} \boldsymbol{\gamma}_t^\top \cdot \boldsymbol{p}^*$$

It follows immediately that $L(\boldsymbol{p}^*) = \sum_{t=1}^{T} L_t(\boldsymbol{p}^*)$. □

Proof of Lemma 4.2.1:. Note that the following two properties are satisfied by the update rule (A.11):

(i). If $\|\boldsymbol{p}_t\|_\infty \leq q$, then we must have $\|\boldsymbol{p}_{t+1}\|_\infty \leq q+1$ by noting that for each i, the i-th component of $\boldsymbol{p}_t$, denoted as $p_{t,i}$, is nonnegative and $g_i(\cdot, \boldsymbol{\theta}_t)$ is normalized within $[0,1]$.

(ii). If there exists i such that $p_{t,i} > q$, then we must have $p_{t+1,i} < p_{t,i}$. Specifically, when $p_{t,i} > q$, we must have that $g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) = 0$, otherwise we would have that

$$f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \leq f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - p_{t,i} \cdot g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) < 0$$

which contradicts the definition of $\tilde{\boldsymbol{x}}_t$ in Algorithm 3 since we could always select $\boldsymbol{x}_t = \boldsymbol{0}$ to obtain a zero objective value as per Assumption 4.1. Then from (A.11), it holds that $p_{t+1,i} < p_{t,i}$.

Starting from $\boldsymbol{p}_1 = \boldsymbol{0}$ and iteratively applying the above two property to control the increase of $\boldsymbol{p}_t$ from $t = 1$ to T, we obtain that for the first time that one component of $\boldsymbol{p}_t$ exceeds the threshold q, it is upper bounded by $q+1$ and this component will continue to decrease until it falls below the threshold q. Thus, we have $\|\boldsymbol{p}_t\|_\infty \leq q+1$ with probability 1 for each t. □

Proof of Theorem 4.3:. In IGD($\boldsymbol{\gamma}$), the true action $\boldsymbol{x}_t$ taken by the decision maker differs from the virtual action $\tilde{\boldsymbol{x}}_t$ if and only if $\boldsymbol{c}_t$ cannot fully satisfy $\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)$. Thus, we have that

$$f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - f(\boldsymbol{x}_t; \boldsymbol{\theta}_t) \leq f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \cdot \mathbb{I}\left\{\exists i : c_{t,i} < g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right\}$$

where $c_{t,i}$ denotes the i-th component of $\boldsymbol{c}_t$ and $\mathbb{I}\{\cdot\}$ denotes the indicator function. Moreover, we know

$$\mathbb{I}\left\{\exists i : c_{t,i} < g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right\} \leq \sum_{i=1}^{m} \mathbb{I}\left\{\sum_{j=1}^{t} g_i(\tilde{\boldsymbol{x}}_j; \boldsymbol{\theta}_j) > c_i\right\}.$$

Recall that the maximum reward generated by consuming per unit of budget of each constraint is upper bounded by q. We have

$$f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \cdot \mathbb{I}\left\{\exists i : c_{t,i} < g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right\} \leq q \sum_{i=1}^{m} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}) \cdot \mathbb{I}\left\{\sum_{j=1}^{t} g_i(\tilde{\boldsymbol{x}}_j; \boldsymbol{\theta}_j) > c_i\right\}$$

From the fact that $\boldsymbol{g}(\cdot; \boldsymbol{\theta}_t) \in [0,1]^m$,

$$\begin{aligned}\sum_{t=1}^{T} f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \sum_{t=1}^{T} f(\boldsymbol{x}_t; \boldsymbol{\theta}_t) &\leq q \cdot \sum_{i=1}^{m} \sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \cdot \mathbb{I}\left\{\sum_{j=1}^{t} g_i(\tilde{\boldsymbol{x}}_j; \boldsymbol{\theta}_j) > c_i\right\} \\ &\leq q \cdot \sum_{i=1}^{m} \left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1)\right]^+\end{aligned}$$

which related the total collected reward by the true action $\{\boldsymbol{x}_t\}_{t=1}^T$ and the virtual action $\{\tilde{\boldsymbol{x}}_t\}_{t=1}^T$. Further from Proposition 4.2, we have that

$$\begin{aligned}\mathrm{Reg}_T(\pi) \leq \min_{\boldsymbol{p} \geq 0} \sum_{t=1}^{T} L_t(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^{T} f(\boldsymbol{x}_t; \boldsymbol{\theta}_t)\right] \leq \underbrace{\sum_{t=1}^{T} \min_{\boldsymbol{p} \geq 0} L_t(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^{T} f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right]}_{\mathrm{I}} \\ \underbrace{+ q \cdot \mathbb{E}\left[\sum_{i=1}^{m} \left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1)\right]^+\right]}_{\mathrm{II}}\end{aligned}$$

We then bound the term I and term II separately to derive our regret bound.

Bound I: Note that for each t, the distribution of $\boldsymbol{p}_t$ is independent from the distribution of $\boldsymbol{\theta}_\tau$

for any $\tau \leq t$, then we have that

$$\min_{\boldsymbol{p} \geq \mathbf{0}} L_t(\boldsymbol{p}) \leq \mathbb{E}_{\boldsymbol{p}_t}\left[L_t(\boldsymbol{p}_t)\right] = \mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{\gamma}_t^\top \boldsymbol{p}_t + \mathcal{P}_t h(\boldsymbol{p}_t; \boldsymbol{\theta}_t)\right]$$

where the expectation is taken with respect to the randomness of the dual price $\boldsymbol{p}_t$. Thus, we have

$$\mathrm{I} \leq \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{\gamma}_t^\top \boldsymbol{p}_t + \mathcal{P}_t\left\{h(\boldsymbol{p}_t; \boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right\}\right]$$

From the definition of $\tilde{\boldsymbol{x}}_t$, we get that $h(\boldsymbol{p}_t; \boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) = -\boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)$, which implies that

$$\mathrm{I} \leq \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{p}_t^\top \cdot \left(\boldsymbol{\gamma}_t - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right)\right]$$

Note that from the update rule (A.11), we have that

$$\|\boldsymbol{p}_{t+1}\|_2^2 \leq \|\boldsymbol{p}_t\|_2^2 + \frac{1}{T} \cdot \|\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{\gamma}_t\|_2^2 - \frac{2}{\sqrt{T}} \cdot \boldsymbol{p}_t^\top \cdot \left(\boldsymbol{\gamma}_t - \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right)$$

which implies that

$$\mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{p}_t^\top \cdot \left(\boldsymbol{\gamma}_t - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right)\right] \leq \frac{\sqrt{T}}{2} \cdot \left(\mathbb{E}[\|\boldsymbol{p}_t\|_2^2] - \mathbb{E}[\|\boldsymbol{p}_{t+1}\|_2^2]\right) + \frac{m}{2\sqrt{T}}$$

Thus, it holds that

$$\mathrm{I} \leq \frac{m\sqrt{T}}{2} \tag{C.1}$$

Bound II: Note that from the update rule (A.11), we have that

$$\sqrt{T} \cdot \boldsymbol{p}_{t+1} \geq \sqrt{T} \cdot \boldsymbol{p}_t + \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{\gamma}_t$$

which implies that

$$\sum_{t=1}^{T} \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{c} \leq \sum_{t=1}^{T} \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \sum_{t=1}^{T} \boldsymbol{\gamma}_t \leq \sqrt{T} \cdot \boldsymbol{p}_{T+1}$$

Thus, it holds that

$$\text{II} = q \cdot \mathbb{E}\left[\sum_{i=1}^{m}\left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1)\right]^+\right] \leq mq(q+1) \cdot \sqrt{T} + qm \tag{C.2}$$

We obtain the $O(\sqrt{T})$ regret bound immediately by combining (C.1) and (C.2). □

Proof of Theorem 4.4:. We construct the following distribution $\mathcal{P}$ over the parameter θ. Suppose that there is a single resource. Denote the support set of $\mathcal{P}$ as $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(j)}, \ldots\}$ and for each $j = 1, 2, \ldots$, we have that

$$\mathcal{P}(\theta = \theta^{(j)}) = \frac{1}{2} \cdot \frac{1}{2^j} \text{ and } \mathcal{P}(\theta = \theta^{(0)}) = \frac{1}{2}$$

Moreover, we have the following condition over the support set $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(j)}, \ldots\}$ of $\mathcal{P}$. Denote sequences of constants $\{a_j\}$, $\{b_j\}$, where

$$a_j = 1 + \frac{2j+1}{2j(j+1)}, \; b_j = 1 + 1/(j+1)$$

for $j = 1, 2, \ldots$. Clearly, we have $1 \leq \cdots \leq a_{j+1} < b_j < a_j \cdots \leq 2$. The reward function satisfies

$$0 \leq \min_{0<x\leq 1} \frac{f(x;\theta^{(0)})}{x} \leq \max_{0<x\leq 1} \frac{f(x;\theta^{(0)})}{x} \leq 1 \text{ and } b_j \leq \min_{0<x\leq 1} \frac{f(x;\theta^{(j)})}{x} \leq \max_{0<x\leq 1} \frac{f(x;\theta^{(j)})}{x} \leq a_j, \; \forall j = 1, 2, \ldots$$

There is a single resource and the resource consumption function satisfies $g(x;\theta) = x$ for each θ. We are now ready to construct the example and show the lower bound.

Suppose there is a single resource with an initial capacity $\frac{T}{2}$, where T is the total time periods.

For a fixed T, there must exists an integer k such that

$$\frac{1}{2\sqrt{T}} \leq \frac{1}{2^{k+2}} \leq \frac{1}{\sqrt{T}}$$

Now, we assume without loss of generality that T is set such that

$$\frac{1}{2^{k+2}} = \frac{1}{\sqrt{T}}$$

We now divide the support set $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(j)}, \ldots\}$ into three subsets:

$$A_1 = \{\theta^{(j)} : 1 \leq j < k\},\ A_2 = \{\theta^{(k)}\} \text{ and } A_3 = \{\theta^{(0)}, \theta^{(j)} : j \geqslant k+1\}.$$

Clearly, at each time period t, we have that

$$P(\theta_t \in A_1) = \frac{1}{2} + \frac{1}{2^{k+1}}, P(\theta_t \in A_2) = \frac{1}{2^{k+1}} \text{ and } P(\theta_t \in A_3) = \frac{1}{2} - \frac{2}{2^{k+1}}.$$

Denote $\varepsilon = \frac{1}{2^{k+2}}$. We call that we encounter a type q request whenever the event $\{\theta_t \in A_q\}$ happens, where $q = 1, 2, 3$. We now denote Z_q^t as the number type q requests in the first t periods. We next introduce the event:

$$\begin{aligned}
\mathcal{H}_1 &= \left\{\frac{1}{\epsilon} \leq Z_2^T \leq \min\{2Z_2^{T/2}, \frac{2}{\epsilon}\}\right\} = \left\{\frac{1}{2}\mathbb{E}[Z_2^T] \leq Z_2^T \leq \min\{2Z_2^{T/2}, \mathbb{E}[Z_2^T]\}\right\} \\
\mathcal{H}_2 &= \left\{Z_1^T \geq \frac{T}{2} + \frac{2}{\epsilon}\right\} = \left\{Z_1^T \geq \mathbb{E}[Z_1^T] + \frac{6}{\epsilon}\right\} \\
\mathcal{H}_3 &= \left\{Z_1^T \leq \frac{T}{2} - \frac{4}{\epsilon}\right\} = \left\{Z_1^T \leq \mathbb{E}[Z_1^T]\right\}
\end{aligned}$$

It is obvious that on event $\mathcal{H}_2$, particularly on the event $\mathcal{H}_2 \cap \mathcal{H}_1$, all the resources will be consumed by type 1 requests by the offline optimum. That is, at each period t, the offline optimum will set the decision variable x_t non-zero only when the event $\{\theta_t \in A_1\}$ happens. Similarly, on

event $\mathcal{H}_1 \cap \mathcal{H}_3$, the offline optimum will set the decision variable $x_t = 1$ whenever the event $\{\theta_t \in A_1\}$ or $\{\theta_t \in A_2\}$ happens. Following the same argument as Arlloto and Gurvich (2019), we know that there exists a constant $\alpha_1 > 0$, independent of ϵ, such that

$$P(\mathcal{H}_1 \cap \mathcal{H}_2) \geq \alpha_1 \text{ and } P(\mathcal{H}_1 \cap \mathcal{H}_3) \geq \alpha_1$$

Next, we consider the difference between the offline optimum and the dynamic programming policy. We denote $S_2^{T/2}$ as the amount of resource consumed by type 2 request during the first $T/2$ periods, and we consider the event

$$\mathcal{H}_4 = \left\{ S_2^{T/2} \geq \frac{Z_2^T}{4} \right\}$$

Now, on the event $\mathcal{H}_4^c \cap \mathcal{H}_1 \cap \mathcal{H}_3$, the offline optimum will set $x_t = 1$ for all type 1 and type 2 request. On the other hand, the optimal online policy consumes at most $Z_2^T/4$ amount of resource using type 2 requests during the first $T/2$ periods. Note that as induced by the event $\mathcal{H}_1$, at most $Z_2^T/2$ number of type 2 request can arrive during the last $T/2$ periods. Thus, for the optimal online policy, type 2 request can consume at most $\frac{3}{4} \cdot Z_2^T$ amount of resource, while the offline optimum consume Z_2^T amount of resource with type 2 request. Further note that the reward/size for type 2 request is at least b_k and the reward/size for type 3 request is at most a_{k+1}. This will incur a regret at least

$$\mathbb{E}[\text{Regret}] \geq \frac{b_k - a_{k+1}}{4} \cdot \mathbb{E}[Z_2^T \cdot 1\{\mathcal{H}_4^c \cap \mathcal{H}_1 \cap \mathcal{H}_3\}] \geq \frac{b_k - a_{k+1}}{4\epsilon} \cdot P(\mathcal{H}_4^c \cap \mathcal{H}_1 \cap \mathcal{H}_3)$$

For each arrival sample path that falls in the set $\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_3$, we can find another sample path in $\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_2$ by keeping the first $T/2$ arrivals unchanged, and for the last $T/2$ periods, replacing at most $6/\epsilon$ type 3 requests with type 1 requests. We denote this resulting set of sample path as $\mathcal{L}$. Note that for each period, the arrival probability for type 1 request and type 3 request are both

bounded away from 0. We know that there exists a constant $\alpha_2 > 0$, independent of ϵ, such that

$$P(\mathcal{L}) \geq \alpha_2 \cdot P(\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_3).$$

Moreover, for each sample path in the set $\mathcal{L}$, since it is on the event $\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_2$, we know that the offline optimum will consume all the resource with type 1 request, while the optimal online policy consumes at least $\frac{Z_2^T}{4}$ amount of resource with type 2 requests. Note that the reward/size for type 1 is at least b_{k-1}, while the reward/size for type 2 is at most a_k. This will incur a regret at least

$$\mathbb{E}[\text{regret}] \geq \frac{b_{k-1} - a_k}{4} \cdot \mathbb{E}[Z_2^T \cdot 1\{\mathcal{L}\}] \geq \frac{b_{k-1} - a_k}{4\epsilon} \cdot P(\mathcal{L}) \geq \alpha_2 \cdot \frac{b_{k-1} - a_k}{4\epsilon} \cdot P(\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_3).$$

Since

$$P(\mathcal{H}_1 \cap \mathcal{H}_3) = P(\mathcal{H}_4^c \cap \mathcal{H}_1 \cap \mathcal{H}_3) + P(\mathcal{H}_4 \cap \mathcal{H}_1 \cap \mathcal{H}_3) \geq \alpha_1,$$

we know that there exists a constant $\alpha_3 > 0$, independent of ϵ, such that

$$\mathbb{E}[\text{regret}] \geq \frac{\alpha_3}{\epsilon} \cdot \min\{b_{k-1} - a_k, b_k - a_{k+1}\}.$$

Recall that $\epsilon = 1/\sqrt{T}$ and

$$\min\{b_{k-1} - a_k, b_k - a_{k+1}\} = \frac{1}{(k+1)(k+2)}$$

and that $\frac{1}{2^{k+2}} = \frac{1}{\sqrt{T}}$, i.e., $k + 2 = \frac{\log T}{2}$. We have

$$\mathbb{E}[\text{regret}] \geq \frac{\alpha_3}{4} \cdot \frac{\sqrt{T}}{(\log T)^2}$$

which completes our proof. □

Proofs of Section 4.3

Proof of Theorem 4.5:. It follows directly from Theorem 4.4 that for any policy π, we have $\mathrm{Reg}_T(\pi) \geq \Omega(\sqrt{T})$. Thus, it is enough to consider the $\Omega(W_T)$ part in the lower bound. We consider the following estimated problem, where the true coefficients in (PCP) are replaced by the estimates:

$$\begin{aligned} \max \ & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \\ \text{s.t.} \ & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, \ldots, T. \end{aligned} \tag{C.3}$$

where $c = \frac{T}{2}$ and the prior estimate $\hat{\mathcal{P}}_t$ is simply a one-point distribution for each t. Now we consider the following two possible true problems, the distributions of which are all one-point distributions and belong to the set Ξ_P with variation budget W_T:

$$\begin{aligned} \max \ & x_1 + \ldots + x_c + \left(1 + \frac{W_T}{T}\right) x_{c+1} + \ldots + \left(1 + \frac{W_T}{T}\right) x_T \\ \text{s.t.} \ & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, \ldots, T. \end{aligned} \tag{C.4}$$

$$\begin{aligned} \max \ & x_1 + \ldots + x_c + \left(1 - \frac{W_T}{T}\right) x_{c+1} + \ldots + \left(1 - \frac{W_T}{T}\right) x_T \\ \text{s.t.} \ & x_1 + \ldots + x_c + x_{c+1} + \ldots + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, \ldots, T. \end{aligned} \tag{C.5}$$

where $c = \frac{T}{2}$. Denote $x_t^1(\pi)$ as the decision of any policy π at period t for scenario (C.4) and denote $x_t^2(\pi)$ as the decision of policy π at period t for scenario (C.5). Further define $T_1(\pi)$ (resp. $T_2(\pi)$) as the expected capacity consumption of policy π on scenario (C.4) (resp. scenario (C.5))

during the first $\frac{T}{2}$ time periods:

$$T_1(\pi) = \mathbb{E}\left[\sum_{t=1}^{\frac{T}{2}} x_t^1(\pi)\right] \quad \text{and} \quad T_2(\pi) = \mathbb{E}\left[\sum_{t=1}^{\frac{T}{2}} x_t^2(\pi)\right]$$

Then, we have that

$$R_T^1(\pi) = \frac{T + W_T}{2} - \frac{W_T}{T} \cdot T_1(\pi) \quad \text{and} \quad R_T^2(\pi) = \frac{T - W_T}{2} + \frac{W_T}{T} \cdot T_2(\pi)$$

where $R_T^1(\pi)$ (resp. $R_T^2(\pi)$) denotes the expected reward collected by policy π on scenario (C.4) (resp. scenario (C.5)). Thus, the regret of policy π on scenario (C.4) and (C.5) are $\frac{W_T}{T} \cdot T_1(\pi)$ and $W_T - \frac{W_T}{T} \cdot T_2(\pi)$ respectively. Further note that since the implementation of policy π at each time period should be only dependent on the historical information and the coefficients in the estimated problem (C.3), we must have $T_1(\pi) = T_2(\pi)$. Thus, we have that

$$\mathrm{Reg}_T(\pi) \geq \max\left\{\frac{W_T}{T} \cdot T_1(\pi), W_T - \frac{W_T}{T} \cdot T_1(\pi)\right\} \geq \frac{W_T}{2} = \Omega(W_T)$$

which completes our proof. □

Proof of Lemma 4.5.1:. Due to symmetry, it is sufficient to show that for every $\boldsymbol{p}$ such that $\boldsymbol{p} \in \Omega_{\bar{p}}$,

$$L_{Q_2}(\boldsymbol{p}) - L_{Q_1}(\boldsymbol{p}) \leq \max\{1, \bar{p}\} \cdot W(Q_1, Q_2).$$

Denote $Q_{1,2}^*$ as the optimal coupling of the distribution Q_1 and Q_2, i.e., the optimal solution to (4.8), and denote

$$\boldsymbol{x}^*(\boldsymbol{\theta}) = \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}}\left\{f(\boldsymbol{x}; \boldsymbol{\theta}) - \sum_{i=1}^{m} p_i \cdot g_i(\boldsymbol{x}; \boldsymbol{\theta})\right\}$$

Then for each $\theta_1 \in \Theta$, we define

$$\hat{x}(\theta_1) = \int_{\theta_2 \in \Theta} x^*(\theta_2) \frac{dQ^*_{1,2}(\theta_1, \theta_2)}{dQ_1(\theta_1)}$$

where $\frac{dQ^*_{1,2}(\theta_1,\theta_2)}{dQ_1(\theta_1)}$ is the Radon–Nikodym derivative of $Q^*_{1,2}$ with respect Q_1 and it can be interpreted as the conditional distribution of θ_2 given θ_1. Note that from the definition of $Q^*_{1,2}$, we have that $\int_{\theta_2 \in \Theta} \frac{dQ^*_{1,2}(\theta_1,\theta_2)}{dQ_1(\theta_1)} = 1$. Thus, $\hat{x}(\theta_1)$ is actually a convex combination of $\{x^*(\theta_2)\}_{\forall \theta_2 \in \Theta}$. Moreover, from the concavity of $f(\cdot; \theta_1)$ and the convexity of $g_i(\cdot; \theta_1)$ for each i, we have that

$$f(\hat{x}(\theta_1); \theta_1) \geq \int_{\theta_2 \in \Theta} f(x^*(\theta_2); \theta_1) \cdot \frac{dQ^*_{1,2}(\theta_1, \theta_2)}{dQ_1(\theta_1)}$$

and

$$g_i(\hat{x}(\theta_1); \theta_1) \leq \int_{\theta_2 \in \Theta} g_i(x^*(\theta_2); \theta_1) \cdot \frac{dQ^*_{1,2}(\theta_1, \theta_2)}{dQ_1(\theta_1)}$$

Thus, we have that

$$\begin{aligned}
L_{Q_1}(p) &= \int_{\theta_1 \in \Theta} \max_{x \in X} \left\{ f(x; \theta_1) - \sum_{i=1}^{m} p_i \cdot g_i(x; \theta_1) \right\} dQ_1(\theta_1) \\
&\geq \int_{\theta_1 \in \Theta} \left\{ f(\hat{x}(\theta_1); \theta_1) - \sum_{i=1}^{m} p_i \cdot g_i(\hat{x}(\theta_1); \theta_1) \right\} dQ_1(\theta_1) \\
&\geq \int_{\theta_1 \in \Theta} \int_{\theta_2 \in \Theta} \left\{ f(x^*(\theta_2); \theta_1) - \sum_{i=1}^{m} p_i \cdot g_i(x^*(\theta_2); \theta_1) \right\} dQ^*_{1,2}(\theta_1, \theta_2)
\end{aligned}$$

Also, note that for any $\theta_1, \theta_2 \in \Theta$, it holds that

$$f(x^*(\theta_2); \theta_1) - \sum_{i=1}^{m} p_i \cdot g_i(x^*(\theta_2); \theta_1) \geq f(x^*(\theta_2); \theta_2) - \sum_{i=1}^{m} p_i \cdot g_i(x^*(\theta_2); \theta_2) - \max\{1, \bar{p}\} \cdot (m+1) \rho(\theta_1, \theta_2)$$

which follows the definition of $\rho(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in (4.2). Thus, we get that

$$\begin{aligned}
L_{Q_1}(\boldsymbol{p}) \geq & \int_{\boldsymbol{\theta}_1 \in \Theta} \int_{\boldsymbol{\theta}_2 \in \Theta} \left\{ f(\boldsymbol{x}^*(\boldsymbol{\theta}_2); \boldsymbol{\theta}_2) - \sum_{i=1}^{m} p_i \cdot g_i(\boldsymbol{x}^*(\boldsymbol{\theta}_2); \boldsymbol{\theta}_2) \right\} dQ^*_{1,2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
& - \max\{1, \bar{p}\} \cdot (m+1) \int_{\boldsymbol{\theta}_1 \in \Theta} \int_{\boldsymbol{\theta}_2 \in \Theta} \rho(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) dQ^*_{1,2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
= & \int_{\boldsymbol{\theta}_2 \in \Theta} \left\{ f(\boldsymbol{x}^*(\boldsymbol{\theta}_2); \boldsymbol{\theta}_2) - \sum_{i=1}^{m} p_i \cdot g_i(\boldsymbol{x}^*(\boldsymbol{\theta}_2); \boldsymbol{\theta}_2) \right\} dQ_2(\boldsymbol{\theta}_2) - \max\{1, \bar{p}\} \cdot (m+1) \mathcal{W}(Q_1, Q_2) \\
= & L_{Q_2}(\boldsymbol{p}) - \max\{1, \bar{p}\} \cdot (m+1) \mathcal{W}(Q_1, Q_2)
\end{aligned}$$

where the first equality holds by noting that $\int_{\boldsymbol{\theta}_1 \in \Theta} dQ^*_{1,2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = dQ_2(\boldsymbol{\theta}_2)$. □

As a remark, we note that the proof of Lemma 4.5.1 will still go through even when the concavity and convexity of $f(\cdot; \boldsymbol{\theta})$ and $g_i(\cdot; \boldsymbol{\theta})$ do not hold. To see this, we use F to denote a distribution over the action set $\mathcal{X}$ and accordingly,

$$\hat{f}(F; \boldsymbol{\theta}) = \int_{\boldsymbol{x} \in \mathcal{X}} \hat{f}(\boldsymbol{x}; \boldsymbol{\theta}) dF(\boldsymbol{x}) \text{ and } \hat{g}_i(F; \boldsymbol{\theta}) = \int_{\boldsymbol{x} \in \mathcal{X}} g_i(\boldsymbol{x}; \boldsymbol{\theta}) dF(\boldsymbol{x}).$$

Then, we denote

$$\hat{h}(\boldsymbol{p}; \boldsymbol{\theta}) := \max_{F} \left\{ \hat{f}(F; \boldsymbol{\theta}) - \boldsymbol{p}^\top \hat{\boldsymbol{g}}(F; \boldsymbol{\theta}) \right\}, \quad \hat{L}_Q(\boldsymbol{p}) = Q\hat{h}(\boldsymbol{p}; \boldsymbol{\theta})$$

and

$$\hat{\rho}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sup_{F} \|(\hat{f}(F; \boldsymbol{\theta}), \hat{\boldsymbol{g}}(F; \boldsymbol{\theta})) - (\hat{f}(F; \boldsymbol{\theta}'), \hat{\boldsymbol{g}}(F; \boldsymbol{\theta}'))\|_\infty, \ \hat{\mathcal{W}}(Q_1, Q_2) := \inf_{Q_{1,2} \in \mathcal{J}(Q_1, Q_2)} \int \hat{\rho}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) dQ_{1,2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Now that $\hat{f}(F; \boldsymbol{\theta})$ and $\hat{\boldsymbol{g}}(F; \boldsymbol{\theta})$ can be regarded as linear functions of $(dF(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathcal{X})$, which fully characterizes the distribution F, we can apply the same procedure as the proof of Lemma 4.5.1 to show that

$$\sup_{\boldsymbol{p} \in \Omega_{\bar{p}}} \left| \hat{L}_{Q_1}(\boldsymbol{p}) - \hat{L}_{Q_2}(\boldsymbol{p}) \right| \leq \max\{1, \bar{p}\} \cdot (m+1) \hat{\mathcal{W}}(Q_1, Q_2).$$

On the other hand, note that

$$\hat{h}(\boldsymbol{p};\boldsymbol{\theta}) \coloneqq \max_{F}\left\{\hat{f}(F;\boldsymbol{\theta}) - \boldsymbol{p}^{\top}\hat{\boldsymbol{g}}(F;\boldsymbol{\theta})\right\} = \max_{\boldsymbol{x}\in\mathcal{X}}\left\{f(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{p}^{\top}\boldsymbol{g}(\boldsymbol{x};\boldsymbol{\theta})\right\} = h(\boldsymbol{p};\boldsymbol{\theta})$$

and $\hat{\rho}(\boldsymbol{\theta},\boldsymbol{\theta}') = \rho(\boldsymbol{\theta},\boldsymbol{\theta}')$. We know that

$$\left|\hat{L}_{Q_1}(\boldsymbol{p}) - \hat{L}_{Q_2}(\boldsymbol{p})\right| = \left|L_{Q_1}(\boldsymbol{p}) - L_{Q_2}(\boldsymbol{p})\right| \text{ and } \hat{\mathcal{W}}(Q_1,Q_2) = \mathcal{W}(Q_1,Q_2).$$

Therefore, (4.12) can be proved to hold for general $f(\cdot;\boldsymbol{\theta})$ and $g_i(\cdot;\boldsymbol{\theta})$ without any convexity or concavity structure.

Proof of Lemma 4.5.2:. Note that the following two property is satisfied by the update rule (4.11):

(i). If $\|\boldsymbol{p}_t\|_\infty \le q$, then we must have $\|\boldsymbol{p}_{t+1}\|_\infty \le q+1$ by noting that for each i, the i-th component of $\boldsymbol{p}_t$, denoted as $p_{t,i}$, is nonnegative and $g_i(\cdot,\boldsymbol{\theta}_t)$ is normalized within $[0,1]$.

(ii). If there exists i such that $p_{t,i} > q$, then we must have $p_{t+1,i} < p_{t,i}$. Specifically, when $p_{t,i} > q$, we must have that $g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) = 0$, otherwise we would have that

$$f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - \boldsymbol{p}_t^{\top}\cdot\boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) \le f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - p_{t,i}\cdot g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) < 0$$

which contradicts the definition of $\tilde{\boldsymbol{x}}_t$ in IGD($\hat{\boldsymbol{\gamma}}$) since we could always select $\tilde{\boldsymbol{x}}_t = \boldsymbol{0}$ to obtain 0 in the objective value. Then from the non-negativity of $\hat{\boldsymbol{c}}_t$, it holds that $p_{t+1,i} < p_{t,i}$ in (4.11).

Starting from $\boldsymbol{p}_1 = 0$ and iteratively applying the above two property to control the increase of $\boldsymbol{p}_t$ from $t=1$ to T, we obtain that for the first time that one component of $\boldsymbol{p}_t$ exceeds the threshold q, it is upper bounded by $q+1$ and this component will continue to decrease until it falls below the threshold q. Thus, it is obvious that we have $\|\boldsymbol{p}_t\|_\infty \le q+1$ with probability 1 for each t. □

Similar to case of known distribution, we define the following function $\hat{L}_t(\cdot)$, based on the prior estimate $\hat{\mathcal{P}}_t$.

$$\hat{L}_t(\boldsymbol{p}) \coloneqq \hat{\boldsymbol{\gamma}}_t^\top \boldsymbol{p} + \hat{\mathcal{P}}_t h(\boldsymbol{p}; \boldsymbol{\theta}). \tag{C.6}$$

Then we have the following relation between $\hat{L}(\cdot)$ and $\hat{L}_t(\cdot)$. As its analysis is identical to Proposition 4.2, we omit its proof for simplicity.

Lemma C.0.1. *For each $t = 1, ..., T$, it holds that*

$$\hat{\boldsymbol{p}}^* \in argmin_{\boldsymbol{p} \geq 0} \hat{L}_t(\boldsymbol{p}) \tag{C.7}$$

where $\hat{\boldsymbol{p}}^*$ is defined in (4.9) as the minimizer of the function $\hat{L}(\cdot)$. Moreover, it holds that

$$\hat{L}(\hat{\boldsymbol{p}}^*) = \sum_{t=1}^{T} \hat{L}_t(\hat{\boldsymbol{p}}^*). \tag{C.8}$$

Now we proof Theorem 4.6 and the idea of proof is similar to Theorem 4.3.

Proof of Theorem 4.6:. From the proof of Theorem 4.3, we have

$$\begin{aligned}
\sum_{t=1}^{T} f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \sum_{t=1}^{T} f(\boldsymbol{x}_t; \boldsymbol{\theta}_t) &\leq q \cdot \sum_{i=1}^{m} \sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \cdot \mathbb{I}\left\{ \sum_{j=1}^{t} g_i(\tilde{\boldsymbol{x}}_j; \boldsymbol{\theta}_j) > c_i \right\} \\
&\leq q \cdot \sum_{i=1}^{m} \left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1) \right]^+
\end{aligned}$$

which related the total collected reward by the true action $\{\boldsymbol{x}_t\}_{t=1}^T$ and the virtual action $\{\tilde{\boldsymbol{x}}_t\}_{t=1}^T$ of the algorithm IGD($\hat{\boldsymbol{y}}$). Further from Proposition 4.2, we have that

$$\begin{aligned}\mathrm{Reg}_T(\pi) \leq \min_{\boldsymbol{p}\geq 0} L(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^T f(\boldsymbol{x}_t;\boldsymbol{\theta}_t)\right] \leq \underbrace{\min_{\boldsymbol{p}\geq 0} L(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^T f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)\right]}_{\mathrm{I}} \\ \underbrace{+ q\cdot\mathbb{E}\left[\sum_{i=1}^m \left[\sum_{t=1}^T g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - (c_i-1)\right]^+\right]}_{\mathrm{II}}\end{aligned}$$

We then bound the term I and term II separately to derive our regret bound.

Bound I: Note that $\|\hat{\boldsymbol{p}}^*\|_\infty \leq q$, it follows directly from Lemma 4.5.1 and Lemma C.0.1 that

$$\min_{\boldsymbol{p}\geq \boldsymbol{0}} L(\boldsymbol{p}) \leq L(\hat{\boldsymbol{p}}^*) \leq \hat{L}(\hat{\boldsymbol{p}}^*) + \max\{q,1\}\cdot(m+1)\cdot W_T = \sum_{t=1}^T \hat{L}_t(\hat{\boldsymbol{p}}^*) + \max\{q,1\}\cdot(m+1)\cdot W_T$$

Note that from Lemma 4.5.2, we have that for each t, $\|\boldsymbol{p}_t\|_\infty \leq (q+1)$ with probability 1. Further note that for each t, the distribution of $\boldsymbol{p}_t$ is independent from the distribution of $\boldsymbol{\theta}_t$, then from Lemma 4.5.1 and Lemma C.0.1, we have that

$$\hat{L}_t(\hat{\boldsymbol{p}}^*) = \min_{\boldsymbol{p}\geq \boldsymbol{0}} \hat{L}_t(\boldsymbol{p}) \leq \mathbb{E}_{\boldsymbol{p}_t}\left[\hat{L}_t(\boldsymbol{p}_t)\right] \leq \mathbb{E}_{\boldsymbol{p}_t}\left[\hat{\boldsymbol{y}}_t^\top \boldsymbol{p}_t + \mathcal{P}_t h(\boldsymbol{p}_t;\boldsymbol{\theta}_t)\right] + (q+1)(m+1)\cdot W(\mathcal{P}_t,\hat{\mathcal{P}})$$

where the expectation is taken with respect to the randomness of the dual price $\boldsymbol{p}_t$. Thus, we have

$$\mathrm{I} \leq \sum_{t=1}^T \mathbb{E}_{\boldsymbol{p}_t}\left[\hat{\boldsymbol{y}}_t^\top \boldsymbol{p}_t + \mathcal{P}_t\left\{h(\boldsymbol{p}_t;\boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)\right\}\right] + 2(q+1)(m+1)\cdot W_T.$$

From the definition of $\tilde{\boldsymbol{x}}_t$, we get that $h(\boldsymbol{p}_t;\boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) = -\boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)$, which implies that

$$\mathrm{I} \leq \sum_{t=1}^T \mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{p}_t^\top \cdot \left(\hat{\boldsymbol{y}}_t - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)\right)\right] + 2(q+1)(m+1)\cdot W_T$$

Note that from the update rule (4.11), we have that

$$\|\boldsymbol{p}_{t+1}\|_2^2 \leq \|\boldsymbol{p}_t\|_2^2 + \frac{1}{T} \cdot \|\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \hat{\boldsymbol{y}}_t\|_2^2 - \frac{2}{\sqrt{T}} \cdot \boldsymbol{p}_t^\top \cdot \left(\hat{\boldsymbol{y}}_t - \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right)$$

which implies that

$$\mathbb{E}_{\boldsymbol{p}_t}\left[\boldsymbol{p}_t^\top \cdot \left(\hat{\boldsymbol{y}}_t - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)\right)\right] \leq \frac{\sqrt{T}}{2} \cdot \left(\mathbb{E}[\|\boldsymbol{p}_t\|_2^2] - \mathbb{E}[\|\boldsymbol{p}_{t+1}\|_2^2]\right) + \frac{m}{2\sqrt{T}}$$

Thus, it holds that

$$\mathrm{I} \leq \frac{m\sqrt{T}}{2} + 2(q+1)(m+1) \cdot W_T \tag{C.9}$$

Bound II: Note that from the update rule (4.11), we have that

$$\sqrt{T} \cdot \boldsymbol{p}_{t+1} \geq \sqrt{T} \cdot \boldsymbol{p}_t + \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \hat{\boldsymbol{y}}_t$$

which implies that

$$\sum_{t=1}^{T} \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{c} \leq \sum_{t=1}^{T} \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \sum_{t=1}^{T} \hat{\boldsymbol{y}}_t \leq \sqrt{T} \cdot \boldsymbol{p}_{T+1}$$

Thus, it holds that

$$\mathrm{II} = q \cdot \mathbb{E}\left[\sum_{i=1}^{m}\left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1)\right]^+\right] \leq mq(q+1) \cdot \sqrt{T} + qm \tag{C.10}$$

We obtain the $O(\max\{\sqrt{T}, W_T\})$ regret bound immediately by combining (C.9) and (C.10). □

Proof of Proposition 4.8. We first describe a problem instance. Consider a single resource with initial capacity $c = 3T/4$. The parameter $\boldsymbol{\theta}_t = (r_t, 1)$; the functions $f(x_t, \boldsymbol{\theta}_t) = r_t \cdot x_t$ and $g(x_t, \boldsymbol{\theta}_t) = x_t$, where $x_t \in [0, 1]$ is the decision variable at time period t. Suppose that the prior

estimate $\hat{\mathcal{P}}$ is given by

$$r_t = \begin{cases} 1, & \text{w.p. } \dfrac{3}{4} \\ \text{Unif}[0,1], & \text{w.p. } \dfrac{1}{4} \end{cases}$$

for each t. Here Unif$[0,1]$ denotes a uniform distribution over $[0,1]$. Clearly, the deterministic upper bound under prior estimate $\hat{\mathcal{P}}$ takes a value of $\frac{3T}{4}$. Furthermore, we denote $\mathbb{E}[h_t^\pi(r)]$ as the expectation of the decision variable under the static policy π when the reward of period t is realized as r. Then, we make following claim.

Claim: There exists a $\hat{r} \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$ such that $\sum_{t=1}^{3T/4} \mathbb{E}[h_t^\pi(\hat{r})] \le T/4$.

Otherwise, suppose that for each $r \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$, we have $\sum_{t=1}^{3T/4} \mathbb{E}[h_t^\pi(r)] \ge T/4$. Then, we compute the expected reward gained by the policy π from $r \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$ during the first $3T/4$ periods, where the budget will never be violated. To be specific, note that the event $r \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$ happens with probability $\frac{W_T}{4T}$ at each period t. Comparing with R_T^{UB}, this will cause a gap of at least $\frac{W_T}{T}$ for π to collect a reward $r \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$. Thus, the gap of R_T^{UB} and π on $\hat{\mathcal{P}}$, caused from π obtaining reward from $r \in [1 - \frac{2W_T}{T}, 1 - \frac{W_T}{T}]$ during the first $3T/4$ periods is at least

$$\frac{W_T}{T} \cdot \frac{W_T}{4T} \cdot \frac{T}{4} = C_1 \cdot T^{1/2}$$

which violates the regret upper bound. Thus, the claim is proved.

Denote by $\hat{r}$ the value described in the claim. Now we construct a true distribution $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_T\}$ as follows:

$$r_t = \begin{cases} \hat{r}, & \text{w.p. } \dfrac{3}{4} \\ \text{Unif}[0,1], & \text{w.p. } \dfrac{1}{4} \end{cases}$$

Clearly, the deviation budget is upper bounded by W_T. Then, we compute the total expected reward that policy π can collect on the true distribution $\mathcal{P}$. The expected reward collected by

policy π during the first $3T/4$ periods is at most

$$\frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3T}{4} + \frac{3}{4} \cdot \frac{T}{4} \cdot \hat{r} \leq \frac{9T}{32}$$

where the first term in the LHS denotes the upper bound of the expected reward collected from Unif$[0, 1]$, and the second term in the LHS denotes the upper bound of the expected reward collect from $\hat{r}$, which follows from the claim. Finally, without considering the budget violation and we let π to collect every reward during the last $T/4$ periods. The policy π can collect reward at most

$$\frac{T}{4} \cdot (\frac{1}{2} \cdot \frac{1}{4} + \frac{3}{4}) = \frac{7T}{32}$$

during the last $T/4$ periods. Thus, the policy π can collect at most $T/2$ reward on the true distribution $\mathcal{P}$ during the entire horizon. However, the value of $\mathbb{E}_{\mathcal{P}}[R_T^{\mathrm{UB}}]$ is at least $\frac{3T}{4} \cdot \hat{r} \geq \frac{3T}{4} \cdot (1 - \frac{2W_T}{T}) \geq \frac{9T}{16}$, when $W_T \leq \frac{T}{8}$ which clearly holds since W_T grows sublinearly in T. Thus, the regret of policy π on the true distribution $\mathcal{P}$ is at least $\frac{T}{16}$. □

Discussion on Bid Price Policy. The well-known bid price policy (Talluri and Van Ryzin (1998)) computes a dual optimal solution (bid price) $\hat{\boldsymbol{p}}^*$ based on the prior estimates. Specifically, the policy makes the decision $\boldsymbol{x}_t$ at each period t based on the fixed $\hat{\boldsymbol{p}}^*$ throughout the procedure:

$$\boldsymbol{x}_t \in \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}, \boldsymbol{\theta}_t) - (\hat{\boldsymbol{p}}^*)^\top \cdot \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}_t).$$

When there is no deviation between the true distributions and prior estimates, i.e., $\mathcal{P}_t = \hat{\mathcal{P}}_t$ for each t, Talluri and Van Ryzin (1998) show that bid price policy is asymptotically optimal if each period is repeated sufficiently many times and the capacity is scaled up accordingly. The following example shows that the bid price policy may fail drastically even when the deviations between the true distributions and prior estimates are small. The example follows the same spirit as the lower bound examples in our paper.

Consider the following linear program as the underlying problem (PCP) for the online stochastic optimization problem:

$$\begin{aligned} \max \ & x_1 + ... + x_{2c} + \frac{1}{2}x_{2c+1} + ... + \frac{1}{2}x_T \\ \text{s.t.} \ & x_1 + ... + x_{2c} + x_{2c+1} + ... + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, ..., T. \end{aligned} \tag{C.11}$$

where $c = \frac{T}{3}$ and without loss of generality, we assume c is an integer. Suppose that the prior estimates for the coefficients in the objective function is larger than the true coefficients by 2ϵ for the first $\frac{T}{3}$ time periods and by ϵ for the last $\frac{2T}{3}$ time periods. Then we obtain the following linear program based on the prior estimates.

$$\begin{aligned} \max \ & (1+2\epsilon)x_1 + ... + (1+2\epsilon)x_c + (1+\epsilon)x_{c+1} + ... + (1+\epsilon)x_{2c} + (\frac{1}{2}+\epsilon)x_{2c+1} + ... + (\frac{1}{2}+\epsilon)x_T \\ \text{s.t.} \ & x_1 + ... + x_{2c} + x_{2c+1} + ... + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, ..., T. \end{aligned} \tag{C.12}$$

Obviously, the optimal dual solution for (C.12) can take any value in $(1+\epsilon, 1+2\epsilon)$. When we apply such a bid price policy to the true problem (C.11), the policy will set $x_t = 0$ throughout the horizon as the reward per time period under the true problem (C.11) is no greater than 1. Given the optimal objective value is $\frac{T}{3}$, the bid price policy will incur a regret of $\frac{T}{3}$.

As a remark, we note that the regret bound for our algorithm IGD($\hat{y}$) is upper bounded by $2\epsilon T + \sqrt{T}$ which can be much smaller than $\frac{T}{3}$ for small ϵ.

Proofs of Section 4.4

Proof of Proposition 4.9:. We consider the implementation of any online policy π on the two scenarios (4.13) and (4.14) for $\kappa = 1$, which is replicated as follows for completeness:

$$\begin{aligned} \max \ & x_1 + ... + x_c + 2x_{c+1} + ... + 2x_T && \text{(C.13)} \\ \text{s.t.} \ & x_1 + ... + x_c + x_{c+1} + ... + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, ..., T. \\ \max \ & x_1 + ... + x_c && \text{(C.14)} \\ \text{s.t.} \ & x_1 + ... + x_c + x_{c+1} + ... + x_T \leqslant c \\ & 0 \leqslant x_t \leqslant 1 \ \text{ for } t = 1, ..., T. \end{aligned}$$

where $c = \frac{T}{2}$. Denote $x_t^1(\pi)$ as the decision of policy π at period t for scenario (C.13) and denote $x_t^2(\pi)$ as the decision of policy π at period t for scenario (C.14). Further define $T_1(\pi)$ (resp. $T_2(\pi)$) as the expected capacity consumption of policy π on scenario (C.13) (resp. scenario (C.14)) during the first $\frac{T}{2}$ time periods:

$$T_1(\pi) = \mathbb{E}\left[\sum_{t=1}^{\frac{T}{2}} x_t^1(\pi)\right] \quad \text{and} \quad T_2(\pi) = \mathbb{E}\left[\sum_{t=1}^{\frac{T}{2}} x_t^2(\pi)\right]$$

Then, we have that

$$R_T^1(\pi) = T - T_1(\pi) \quad \text{and} \quad R_T^2(\pi) = T_2(\pi)$$

where $R_T^1(\pi)$ (resp. $R_T^2(\pi)$) denotes the expected reward collected by policy π on scenario (C.13) (resp. scenario (C.14)). Thus, the regret of policy π on scenario (C.13) and (C.14) are $T_1(\pi)$ and $T - T_2(\pi)$ respectively. Further note that since the implementation of policy π at each time period should be independent of the future information, we must have $T_1(\pi) = T_2(\pi)$. Thus, we have

that

$$\mathrm{Reg}_T(\pi) \geq \max\{T_1(\pi), T - T_1(\pi)\} \geq \frac{T}{2} = \Omega(T)$$

which completes our proof. □

Proof of Theorem 4.10:. The proof of the theorem can be directly obtained from Theorem 4.5.

□

We first prove the following lemma, which implies that the dual variable updated in (4.15) is always bounded. It derivation is essentially the same as Lemma 4.2.1, so we omit its proof for simplicity.

Lemma C.0.2. *Under Assumption 4.1, for each $t = 1, 2, \ldots, T$, the dual price vector satisfies $\|\boldsymbol{p}_t\|_\infty \leq q + 1$, where $\boldsymbol{p}_t$ is specified by (4.15) in Algorithm 4 and the constant q is defined in Assumption 4.1 (c).*

Now we proceed to prove Theorem 4.11.

Proof of Theorem 4.11:. From the proof of Theorem 4.3, we have

$$\sum_{t=1}^{T} f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \sum_{t=1}^{T} f(\boldsymbol{x}_t; \boldsymbol{\theta}_t) \leq q \cdot \sum_{i=1}^{m} \left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - (c_i - 1) \right]^+$$

which relates the total collected reward by the true action $\{\boldsymbol{x}_t\}_{t=1}^T$ and the virtual action $\{\tilde{\boldsymbol{x}}_t\}_{t=1}^T$. Here $[\cdot]^+$ denotes the positive part function. Furthermore, from Proposition 4.2 and the feasibility,

we have that

$$\mathrm{Reg}_T(\pi) \le \min_{\boldsymbol{p}\ge \mathbf{0}} L(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^{T} f(\boldsymbol{x}_t;\boldsymbol{\theta}_t)\right] \le \underbrace{\min_{\boldsymbol{p}\ge \mathbf{0}} L(\boldsymbol{p}) - \mathbb{E}\left[\sum_{t=1}^{T} f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)\right]}_{\mathrm{I}} + \underbrace{q\cdot \mathbb{E}\left[\sum_{i=1}^{m}\left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - (c_i-1)\right]^+\right]}_{\mathrm{II}}$$

Next, we bound the term I and term II separately to derive our regret bound.

Bound I: We first define the following function $\bar{L}(\cdot)$:

$$\bar{L}(\boldsymbol{p}) \coloneqq \frac{1}{T}\boldsymbol{p}^\top \boldsymbol{c} + \bar{\mathcal{P}}_T h(\boldsymbol{p};\boldsymbol{\theta})$$

Note that $\hat{\mathcal{P}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathcal{P}_t$, it holds that $L(\boldsymbol{p}) = T\cdot\bar{L}(\boldsymbol{p})$ for any $\boldsymbol{p}$. From Lemma C.0.2, we know that for each t, $\|\boldsymbol{p}_t\|_\infty \le q+1$ with probability 1. In addition, for each t, the distribution of $\boldsymbol{p}_t$ is independent from the distribution of $\boldsymbol{\theta}_t$, then from Lemma 4.5.1, we have that

$$\min_{\boldsymbol{p}\ge\mathbf{0}} \bar{L}(\boldsymbol{p}) \le \mathbb{E}_{\boldsymbol{p}_t}\left[\bar{L}(\boldsymbol{p}_t)\right] \le \mathbb{E}_{\boldsymbol{p}_t}\left[\frac{1}{T}\boldsymbol{p}_t^\top \boldsymbol{c} + \mathcal{P}_t h(\boldsymbol{p}_t;\boldsymbol{\theta}_t)\right] + (q+1)(m+1)\cdot\mathcal{W}(\mathcal{P}_t,\bar{\mathcal{P}}_T), \tag{C.15}$$

where the expectation is taken with respect to $\boldsymbol{p}_t$ in a random realization of the algorithm. Thus, we have the first term

$$\mathrm{I} \le \sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{p}_t}\left[\frac{1}{T}\boldsymbol{c}^\top\boldsymbol{p}_t + \mathcal{P}_t\left\{h(\boldsymbol{p}_t;\boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t)\right\}\right] + (q+1)(m+1)\cdot\mathcal{W}(\mathcal{P}_t,\bar{\mathcal{P}}_T)$$

which comes from combining (C.15) with the relation $L(\boldsymbol{p}) = T \cdot \bar{L}(\boldsymbol{p})$. By the definition of $\tilde{\boldsymbol{x}}_t$, $h(\boldsymbol{p}_t; \boldsymbol{\theta}_t) - f(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) = -\boldsymbol{p}_t^\top \cdot \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t)$, which implies that

$$\mathrm{I} \leq \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{p}_t} \left[\boldsymbol{p}_t^\top \cdot \left(\frac{\boldsymbol{c}}{T} - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \right) \right] + (q+1)(m+1) \cdot \mathcal{W}(\mathcal{P}_t, \bar{\mathcal{P}}_T) \tag{C.16}$$

Note that from the update rule (4.15), we have that

$$\|\boldsymbol{p}_{t+1}\|_2^2 \leq \|\boldsymbol{p}_t\|_2^2 + \frac{1}{T} \cdot \|\boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \frac{\boldsymbol{c}}{T}\|_2^2 - \frac{2}{\sqrt{T}} \cdot \boldsymbol{p}_t^\top \cdot \left(\frac{\boldsymbol{c}}{T} - \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \right).$$

By taking expectation with respect to both sides,

$$\mathbb{E}_{\boldsymbol{p}_t} \left[\boldsymbol{p}_t^\top \cdot \left(\frac{\boldsymbol{c}}{T} - \mathcal{P}_t \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) \right) \right] \leq \frac{\sqrt{T}}{2} \cdot \left(\mathbb{E}[\|\boldsymbol{p}_t\|_2^2] - \mathbb{E}[\|\boldsymbol{p}_{t+1}\|_2^2] \right) + \frac{m}{2\sqrt{T}} \tag{C.17}$$

Plugging (C.17) into (C.16), we obtain an upper bound on Term I,

$$\mathrm{I} \leq \frac{m\sqrt{T}}{2} + (q+1)(m+1) \cdot W_T \tag{C.18}$$

Bound II: Note that from the update rule (4.15), we have

$$\sqrt{T} \cdot \boldsymbol{p}_{t+1} \geq \sqrt{T} \cdot \boldsymbol{p}_t + \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \frac{\boldsymbol{c}}{T}$$

Taking a summation with respect to both sides,

$$\sum_{t=1}^{T} \boldsymbol{g}(\tilde{\boldsymbol{x}}_t; \boldsymbol{\theta}_t) - \boldsymbol{c} \leq \sqrt{T} \cdot \boldsymbol{p}_{T+1}$$

Applying Lemma C.0.2 for a bound on $\boldsymbol{p}_{T+1}$, we obtain the upper bound for Term II,

$$\text{II} = q \cdot \mathbb{E}\left[\sum_{i=1}^{m}\left[\sum_{t=1}^{T} g_i(\tilde{\boldsymbol{x}}_t;\boldsymbol{\theta}_t) - (c_i - 1)\right]^+\right] \leq mq(q+1)\cdot\sqrt{T} + qm \tag{C.19}$$

We obtain the desired regret bound by combining (C.18) and (C.19). □

D | Appendix for Chapter 5

Proof of Lemma 5.2.1:. The equations hold due to our definition of deterministic policies. To see this, consider any given deterministic policy $\boldsymbol{\phi} \in \Phi$. It determines a unique allocation $(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D}), \boldsymbol{s}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D})) \in P(\boldsymbol{c}, \boldsymbol{D})$ for every demand realization $\boldsymbol{D}$. Thus $(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D}), \boldsymbol{s}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D}))$ is a feasible solution to (5.16). This implies that

$$f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{D})) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \boldsymbol{d}), D_j) \geq g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{d})$$

and thus

$$F(\mathbf{w}, \boldsymbol{\phi}) = E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] - \sum_{j \in \mathcal{N}} w_j \cdot E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] \geq E_{\tilde{\mathbf{D}}}[g(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})]$$

holds for any $\boldsymbol{\phi} \in \Phi$. Thus,

$$\inf_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi}) \geqslant E_{\tilde{\mathbf{D}}}[g(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})]$$

On the other hand, by the definition of $\boldsymbol{\phi}_{\mathbf{w}}$, we have that

$$g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{d}) = f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}}, \boldsymbol{c}, \boldsymbol{D})) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}}, \boldsymbol{c}, \boldsymbol{d}), D_j)$$

and thus

$$E_{\tilde{\mathbf{D}}}[g(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})] = F(\mathbf{w}, \boldsymbol{\phi}_{\mathbf{w}}) \geq \inf_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi})$$

Therefore, equality (5.17) holds. □

Proof of Theorem 5.3:. When a capacity level $\boldsymbol{c}$ is asymptotically feasible, from (5.8), it is clear that we have

$$\inf_{\lambda\in\chi} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi})\right) \leq 0$$

for each fixed $\mathbf{w} \geq 0$. Thus, it holds that

$$\sup_{\mathbf{w}\geq 0} \inf_{\lambda\in\chi} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi})\right) \leq 0 \tag{D.1}$$

From Lemma 5.2.1 (assuming zero allocation cost), we immediately tell that (5.18) holds.

We now prove the reverse direction. If (5.18) holds, then (D.1) holds from Lemma 5.2.1. We define the set $W = \{\mathbf{w} \geq 0 : \sum_{j\in\mathcal{N}} w_j \leq 1\}$. Clearly, we have that

$$\max_{\mathbf{w}\in W} \inf_{\lambda\in\chi} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi})\right) \leq 0 \tag{D.2}$$

Obviously, the set χ is a convex set. Moreover, note that W is a convex compact set and the objective function in the above problem is linear in $\mathbf{w}$ (resp. λ) when λ is fixed. Then by Sion's minimax theorem (Sion, 1958), we can interchange the order of max and inf on the left-hand side of (D.2). Thus, we have

$$\inf_{\lambda\in\chi} \max_{\mathbf{w}\in W} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda(\boldsymbol{\phi})\right) \leq 0$$

For any $\epsilon > 0$, there exists a randomized policy $\lambda_\epsilon \in \chi$ such that

$$\sup_{\mathbf{w}\in W} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda_\epsilon(\boldsymbol{\phi})\right) \leq \epsilon \tag{D.3}$$

We now claim that λ_ϵ achieves a service level at least $\beta_j - \epsilon$ for each $j \in \mathcal{N}$, i.e.

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)]d\lambda_\epsilon(\boldsymbol{\phi}) \geq \beta_j - \epsilon. \tag{D.4}$$

Otherwise, suppose there exists a $j' \in \mathcal{N}$ such that (D.4) does not hold. Then we define $\hat{\mathbf{w}} \in W$ such that $\hat{w}_{j'} = 1$ and $\hat{w}_j = 0$ for all $j \neq j'$. It is clear that

$$\begin{aligned}\epsilon &< \hat{w}_{j'} \cdot \left(\beta_{j'} - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_{j'}(s_{j'}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_{j'})]d\lambda_\epsilon(\boldsymbol{\phi})\right) \\ &\leq \sup_{\mathbf{w}\in W_\epsilon} \sum_{j\in\mathcal{N}} w_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)]d\lambda_\epsilon(\boldsymbol{\phi})\right) \leq \epsilon\end{aligned}$$

where the last inequality follows from (D.3). This is a contradiction. Thus, we prove that $\boldsymbol{c}$ is asymptotically feasible, which completes our proof. □

Proof of Theorem 5.5

We first prove the following lemma, which will lead to our final result. Its proof is relegated to the Appendix D.

Lemma D.0.1. *Under Assumption 5.4, there exists a subset of deterministic policies Φ_W such that for any $\mathbf{w} \geqslant 0$, we have*

$$\inf_{\boldsymbol{\phi}\in\Phi} F(\mathbf{w}, \boldsymbol{\phi}) = F(\mathbf{w}, \boldsymbol{\phi}_{\mathbf{w}}) = \min_{\boldsymbol{\phi}\in\Phi_W} F(\mathbf{w}, \boldsymbol{\phi}). \tag{D.5}$$

Moreover, for any sequence of randomized policies $\{\lambda_k\}_{k\geq 1}$ such that $\lambda_k \in \chi_W := \{\lambda \geq 0 : \int_{\boldsymbol{\phi}\in\Phi_W} d\lambda(\boldsymbol{\phi}) = 1\}$ for each integer k, there exists a policy $\hat{\lambda} \in \chi$ such that

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))]d\hat{\lambda}(\boldsymbol{\phi}) \leq \limsup_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))]d\lambda_k(\boldsymbol{\phi}) \tag{D.6}$$

and

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)]d\hat{\lambda}(\boldsymbol{\phi}) \geq \liminf_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)]d\lambda_k(\boldsymbol{\phi}), \quad \forall j \in \mathcal{N}. \tag{D.7}$$

We are now ready to prove the strong duality between (5.11) and (5.12). We first use weak duality to show that Obj (5.11) $\geqslant$ Obj (5.12). In order to prove the other direction, we construct a sequence of randomized policy λ_k such that the achieved service level is at least $\beta_j - \frac{\hat{C}}{k}$ for each $j \in \mathcal{N}$, for some constant $\hat{C} > 0$. Then, by letting $k \to \infty$, we use Lemma D.0.1 to show that there exists a randomized policy such that the target service level β_j is achieved for each $j \in \mathcal{N}$ and the expected allocation cost is upper bounded by Obj (5.12), which completes our proof.

Proof of Theorem 5.5:. We denote Obj (5.11) as the objective value of (5.11) and denote Obj (5.12) as the objective value of (5.12). When (5.11) is feasible, we get

$$\text{Obj (5.11)} \geqslant \text{Obj (5.12)} \tag{D.8}$$

by weak duality (Shapiro, 2001). In the remaining part of the proof, we assume that Obj (5.12) is finite, and we prove that (5.11) is feasible and Obj (5.11) = Obj (5.12).

Note that

$$\inf_{\lambda\in\chi} L(\mathbf{w},\lambda) = \sum_{j\in\mathcal{N}} w_j \cdot \beta_j + \inf_{\boldsymbol{\phi}\in\Phi} F(\mathbf{w},\boldsymbol{\phi}) = \sum_{j\in\mathcal{N}} w_j \cdot \beta_j + \inf_{\boldsymbol{\phi}\in\Phi_W} F(\mathbf{w},\boldsymbol{\phi}) = \inf_{\lambda\in\chi_W} L(\mathbf{w},\lambda) \tag{D.9}$$

where the second equality holds due to Lemma D.0.1. Then we have

$$\sup_{\mathbf{w}\geq 0} \inf_{\lambda\in\chi_W} L(\mathbf{w},\lambda) = \text{Obj (5.12)}$$

For each integer $k > 1$, we define the set $W_k = \{\mathbf{w} \geq 0 : \sum_{j \in \mathcal{N}} w_j \leq k\}$. Obviously, it holds that

$$\sup_{\mathbf{w} \in W_k} \inf_{\lambda \in \chi_W} L(\mathbf{w}, \lambda) \leq \sup_{\mathbf{w} \geq 0} \inf_{\lambda \in \chi_W} L(\mathbf{w}, \lambda) = \text{Obj (5.12)}$$

By definition, χ_W is a convex set. Moreover, note that W_k is a convex compact set and $L(\mathbf{w}, \lambda)$ is linear in $\mathbf{w}$ (resp. λ) when λ (resp. $\mathbf{w}$) is fixed. Then by Sion's minimax theorem (Sion, 1958), we must have

$$\inf_{\lambda \in \chi_W} \sup_{\mathbf{w} \in W_k} L(\mathbf{w}, \lambda) = \inf_{\lambda \in \chi_W} \max_{\mathbf{w} \in W_k} L(\mathbf{w}, \lambda) = \max_{\mathbf{w} \in W_k} \inf_{\lambda \in \chi_W} L(\mathbf{w}, \lambda) = \sup_{\mathbf{w} \in W_k} \inf_{\lambda \in \chi_W} L(\mathbf{w}, \lambda) \leq \text{Obj (5.12)}$$

Thus, there exists a randomized policy $\lambda_k \in \chi_W$ such that

$$\sup_{\mathbf{w} \in W_k} L(\mathbf{w}, \lambda_k) \leq \inf_{\lambda \in \chi_W} \sup_{\mathbf{w} \in W_k} L(\mathbf{w}, \lambda) + \frac{1}{k} \leq \text{Obj (5.12)} + \frac{1}{k} \tag{D.10}$$

Denote $\hat{C} = \text{Obj (5.12)} + 1$. We now claim that λ_k achieves a service level at least $\beta_j - \frac{\hat{C}}{k}$ for each $j \in \mathcal{N}$, i.e.

$$\int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda_l(\boldsymbol{\phi}) \geq \beta_j - \frac{\hat{C}}{k}. \tag{D.11}$$

Otherwise, suppose there exists a $j_0 \in \mathcal{N}$ such that

$$\hat{C} < k \cdot \left(\beta_{j_0} - \int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[R_{j_0}(s_{j_0}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_{j_0})] d\lambda_k(\boldsymbol{\phi}) \right).$$

Then we define $\hat{\mathbf{w}} \in W_k$ such that $\hat{w}_{j_0} = k$ and $\hat{w}_j = 0$ for all $j \neq j_0$. By construction, we have

$$\begin{aligned}
\hat{C} &< \hat{w}_{j_0} \cdot \left(\beta_{j_0} - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_{j_0}(s_{j_0}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_{j_0})] d\lambda_k(\boldsymbol{\phi})\right) \\
&= \sum_{j\in N} \hat{w}_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda_k(\boldsymbol{\phi})\right) \\
&\leqslant \sum_{j\in N} \hat{w}_j \cdot \left(\beta_j - \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda_k(\boldsymbol{\phi})\right) + E_{\tilde{\mathbf{D}}}[f(y(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] \\
&= L(\hat{\mathbf{w}}, \lambda_k) \leq \sup_{\mathbf{w}\in W_k} L(\mathbf{w}, \lambda_k) \leq \hat{C}
\end{aligned}$$

where the second inequality holds since the allocation cost function f is non-negative and the last inequality follows from (D.10) since $1 \geq \frac{1}{k}$. This is a contradiction.

From Lemma D.0.1, for the sequence $\{\lambda_k\}_{k\geq 1}$, there exists a randomized policy $\hat{\lambda} \in \chi$ such that (D.6) and (D.7) hold. Specifically, for each $j \in \mathcal{N}$, we have

$$\begin{aligned}
\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\hat{\lambda}(\boldsymbol{\phi}) &\geq \liminf_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)] d\lambda_k(\boldsymbol{\phi}) \\
&\geq \liminf_{k\to\infty} \beta_j - \frac{\hat{C}}{k} = \beta_j
\end{aligned}$$

Thus, (5.11) is feasible and $\hat{\lambda}$ is a feasible solution to (5.11). Moreover, from (D.6), we have

$$\begin{aligned}
\text{Obj (5.11)} &\leq \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\hat{\lambda}(\boldsymbol{\phi}) \leq \limsup_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\lambda_k(\boldsymbol{\phi}) \\
&\leq \limsup_{k\to\infty} \sup_{\mathbf{w}\in W_k} L(\mathbf{w}, \lambda_k) \leq \limsup_{k\to\infty} \text{Obj (5.12)} + \frac{1}{k} = \text{Obj (5.12)}
\end{aligned}$$

where the third inequality follows from the fact that

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\lambda_k(\boldsymbol{\phi}) = L(\mathbf{0}, \lambda_k) \leq \sup_{\mathbf{w}\in W_k} L(\mathbf{w}, \lambda_k), \quad \forall k > 1$$

and the last inequality follows from (D.10). Together with the weak duality established in (D.8),

we have Obj (5.11) = Obj (5.12). □

Proof of Lemma D.0.1

We first present the following well-known result, which will be useful for our proof of Lemma D.0.1.

Theorem D.1. *(Banach-Saks theorem) Let* $\{\mathbf{x}^k\}_{k=1}^{\infty}$ *be a bounded sequence in the Hilbert space* $\mathcal{H}$*, then there exists a subsequence* $\{n_k\}_{k=1}^{\infty}$ *of* $\{1, 2, \dots, k, \dots\}$ *and a point* $\mathbf{x} \in \mathcal{H}$ *such that*

$$\frac{1}{k} \cdot \sum_{l=1}^{k} \mathbf{x}^{n_l}$$

converges strongly to $\mathbf{x}$ *as* $k \to \infty$*.*

Now we are ready to prove Lemma D.0.1. The following is an outline of the proof of Lemma D.0.1. We first represent each policy as an element in a Hilbert space and interpret each term in (D.6) and (D.7) as an inner product that defines the metric of the Hilbert space. We then apply the Banach-Saks theorem to prove weak convergence of the running average of a subsequence to a limiting element in the Hilbert space, which directly implies that the cost and the service level of the running average converge to those of the limiting element. Finally, we show that the limiting element can be interpreted back as a policy, which establishes the existence of $\hat{\lambda}$ in Lemma D.0.1.

Proof of Lemma D.0.1:. Note that under Assumption 5.4, the service measure function for each $j \in \mathcal{N}$ is piece-wise linear in s_j. Thus, the objective function of (5.16) is a piece-wise linear function. Then, problem (5.16) can be solved by first enumerating all possible vectors $\mathbf{k} \in K := \{(k_1, k_2, \cdots, k_n) : k_j = 1, 2, \dots, K_j\}$, such that

$$a_{j,k_j} \cdot D_j \leq s_j < a_{j,k_j+1} \cdot D_j,$$

and then for each set of values $\mathbf{k}$ solving

$$\begin{aligned} g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}; \mathbf{k}) := \min_{\boldsymbol{s},\mathbf{y}} \; & f(y) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j, D_j) \\ \text{s.t.} \quad & (\boldsymbol{s}, \mathbf{y}) \in P(\boldsymbol{c}, \boldsymbol{D}) \\ & a_{j,k_j} \cdot D_j \le s_j \le a_{j,k_j+1} \cdot D_j \;\; \forall j \in \mathcal{N}. \end{aligned} \tag{D.12}$$

Note that in the problem (D.12), we replace the constraint $s_j < a_{j,k_j+1} \cdot D_j$ with $s_j \le a_{j,k_j+1} \cdot D_j$. This is because for fixed k_j, we can simply define $R_j(a_{j,k_j+1} \cdot D_j, D_j) = \lim_{s_j \to (a_{j,k_j+1} \cdot D_j)-} R_j(s_j, D_j)$ denoting the left limit when solving (D.12). Then the minimum value of $g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{D}; \mathbf{k})$ over all possible $\mathbf{k} \in K$ is still the same as the optimal value of the original problem (5.16).

By definition of $P(\boldsymbol{c}, \boldsymbol{D})$, the linear program (D.12) can be reformulated as:

$$\begin{aligned} g(\mathbf{w}, \boldsymbol{c}, \boldsymbol{D}; \mathbf{k}) = \min \; & \mathbf{r}_{\mathbf{w},\boldsymbol{d}}^T \hat{\mathbf{y}} \\ \text{s.t.} \; & \mathbf{A}\hat{\mathbf{y}} = \mathbf{v}_{\boldsymbol{d}} \\ & \hat{\mathbf{y}} \ge 0 \end{aligned} \tag{D.13}$$

by choosing the appropriate $\mathbf{r}_{\mathbf{w},\boldsymbol{d}}$, $\mathbf{A}$ and $\mathbf{v}_{\mathbf{d}}$, where $\mathbf{v}_{\boldsymbol{d}}$ is independent of $\mathbf{w}$ and $\mathbf{A}$ is independent of $\mathbf{w}$ and $\boldsymbol{d}$. Although all the coefficients $\mathbf{r}_{\mathbf{w},\boldsymbol{d}}$, $\mathbf{v}_{\boldsymbol{d}}$ and $\mathbf{A}$ should also be dependent on $\mathbf{k}$, we drop the dependency for simplicity of notation. The dependence of $\mathbf{v}_{\boldsymbol{d}}$ on $\boldsymbol{c}$ is also dropped since $\boldsymbol{c}$ is fixed. We now focus on LP (D.13).

Denote $\mathcal{D}$ as the support of the demand distribution. Given $\mathbf{k}$, for each $\mathbf{w} \ge 0$ and $\boldsymbol{d} \in \mathcal{D}$, there could be multiple optimal solutions. However, it is enough for us to only consider one optimal *basic solution* that is determined by a basis $\mathbf{b} \in \mathcal{B}$, where $\mathcal{B}$ is the set of all bases of $\mathbf{A}$. Since $\mathbf{A}$ has a finite size, the total number of all bases, $|\mathcal{B}|$ should be finite. Then for any $\mathbf{w} \ge 0$ and $\boldsymbol{d} \in \mathcal{D}$, an optimal solution of (5.16) is uniquely determined by an element in the finite set $\mathcal{V} = \mathcal{K} \times \mathcal{B}$. For the rest of the proof, we only consider such optimal solutions. Without loss of

generality, we sort the elements in the set $\mathcal{V}$ in a fixed sequence and we will use the order of an element in this sequence to denote this element by abuse of notation.

For each fixed $\mathbf{w} \in W$, we define a deterministic policy $\hat{\boldsymbol{\phi}}_{\mathbf{w}}$ as follows. For any $\boldsymbol{d} \in \mathcal{D}$, let $\hat{\boldsymbol{\phi}}_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{d})$ be the specified optimal solution of (5.16) determined by one element from $\mathcal{V}$. As a direct consequence of this definition, for each $\mathbf{w} \in W$, we have that

$$\hat{\boldsymbol{\phi}}_{\mathbf{w}} \in \operatorname{argmin}_{\boldsymbol{\phi} \in \Phi} F(\mathbf{w}, \boldsymbol{\phi}).$$

We define $\Phi_W = \{\hat{\boldsymbol{\phi}}_{\mathbf{w}} : \forall \mathbf{w} \in W\}$. Then, our proof of (D.5) is finished. In the remaining part of the proof, we prove (D.6) and (D.7).

For each $\boldsymbol{D} \in \mathcal{D}$ and each element $v \in \mathcal{V}$, we further denote $a(v, \boldsymbol{D})$ as the basic solution of (D.13) determined by the element v if feasible, i.e., $a(v, \boldsymbol{D}) \in P(\boldsymbol{c}, \boldsymbol{D})$. If infeasible, we simply denote $a(v, \boldsymbol{D}) = \mathbf{0}$. Then, from definition, for each $\mathbf{w} \in W$ and each $\boldsymbol{D} \in \mathcal{D}$, the allocation $\hat{\boldsymbol{\phi}}_{\mathbf{w}}(\boldsymbol{c}, \boldsymbol{d}) \in \{a(v, \boldsymbol{D})\}_{\forall v \in \mathcal{V}}$.

We now focus on the sequence of randomized policies $\{\lambda_k\}$ such that $\lambda_k \in \chi_W := \{\lambda \geq 0 : \int_{\boldsymbol{\phi} \in \Phi_W} d\lambda(\boldsymbol{\phi}) = 1\}$. From the above argument, we know that for any k and any $\boldsymbol{D} \in \mathcal{D}$, the allocation of λ_k is simply a randomization over $\{a(v, \boldsymbol{D})\}_{\forall v \in \mathcal{V}}$. Then, we define a vector $\boldsymbol{\psi}^k(\boldsymbol{D}) \in \mathbb{R}^{|\mathcal{V}|}$, where $|\mathcal{V}|$ denotes the cardinality of the finite set $\mathcal{V}$, such that the v-th component of $\boldsymbol{\psi}^k(\boldsymbol{D})$, denoted as $\psi_v^k(\boldsymbol{D})$, denotes the probability that the allocation of λ_k equals $a(v, \boldsymbol{D})$, given demand realization $\boldsymbol{D}$. Obviously, we have that

$$\boldsymbol{\psi}^k(\boldsymbol{D}) \in \mathcal{L} := \{\mathbf{x} \in \mathbb{R}^{|\mathcal{V}|} : \mathbf{x} \geq 0, \ \sum_{v=1}^{|\mathcal{V}|} x_v = 1\}, \quad \forall \boldsymbol{D} \in \mathcal{D} \tag{D.14}$$

Following this definition, each randomized policy λ_k is equivalently represented by $\boldsymbol{\psi}^k = (\boldsymbol{\psi}(\boldsymbol{D}), \forall \boldsymbol{D} \in \mathcal{D})$, which is a measurable function mapping the set $\mathcal{D} \subset \mathbb{R}^n$ to the set $\mathcal{L} \subset \mathbb{R}^{|\mathcal{V}|}$. From (D.14), it

is easy to see that

$$\|\psi_v^k\|_{L^2}^2 = \int_{\boldsymbol{D}\in\mathcal{D}} |\psi_v^k(\boldsymbol{D})|^2 d\mu(\boldsymbol{D}) \leq 1, \quad \forall v = 1, \ldots, |\mathcal{V}| \tag{D.15}$$

Here, μ denotes the measure over $\mathcal{D}$ specified by the demand distribution and $\|\cdot\|_{L^2}$ denotes the L^2-norm. Then, for each k and each $v = 1, \ldots, |\mathcal{V}|$, we conclude that $\psi_v^k \in L^2(\mathcal{D}, \mu)$, where $L^2(\mathcal{D}, \mu)$ denotes the L^2 space containing all measurable functions over the set $\mathcal{D}$ with finite L^2-norm.

For each $v = 1, \ldots, |\mathcal{V}|$, we further denote L_v as a copy of the space $L^2(\mathcal{D}, \mu)$, which is a Hilbert space. We then denote $\mathcal{H}$ as the direct sum of the spaces $\{L_v\}_{v=1}^{|\mathcal{V}|}$, i.e.,

$$\mathcal{H} = \bigoplus_{v=1}^{|\mathcal{V}|} L_v := \{\forall \boldsymbol{\varphi} = (\varphi_v, v = 1, \ldots, |\mathcal{V}|) \text{ such that } \varphi_v \in L_v \text{ for each } v\}$$

Clearly, $\mathcal{H}$ is still a Hilbert space, equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ defined as:

$$\langle \boldsymbol{\varphi}^1, \boldsymbol{\varphi}^2 \rangle_{\mathcal{H}} = \sum_{v=1}^{|\mathcal{V}|} \int_{\boldsymbol{D}\in\mathcal{D}} \varphi_v^1(\boldsymbol{D}) \cdot \varphi_v^2(\boldsymbol{D}) d\mu(\boldsymbol{D}), \quad \forall \boldsymbol{\varphi}^1, \boldsymbol{\varphi}^2 \in \mathcal{H}$$

Denote $\|\cdot\|_{\mathcal{H}}$ as the norm on the Hilber space $\mathcal{H}$ induced by the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. For each k, it holds directly that $\boldsymbol{\psi}^k \in \mathcal{H}$ and from (D.15), we have

$$\|\boldsymbol{\psi}^k\|_{\mathcal{H}}^2 = \sum_{v=1}^{|\mathcal{V}|} \int_{\boldsymbol{D}\in\mathcal{D}} |\psi_v^k(\boldsymbol{D})|^2 d\mu(\boldsymbol{D}) \leq |\mathcal{V}| \tag{D.16}$$

We then show that the expected allocation cost and service level of the policy λ_k can be expressed as the inner product in the space $\mathcal{H}$. For each $v \in \mathcal{V}$ and each $\boldsymbol{D} \in \mathcal{D}$, we specify allocation $\mathrm{y}(v, \boldsymbol{D})$ and fulfillment $\boldsymbol{s}(v, \boldsymbol{D})$ such that $a(v, \boldsymbol{D}) = (\mathrm{y}(v, \boldsymbol{D}), \boldsymbol{s}(v, D))$. Then, we define

$$\boldsymbol{\eta}^{(0)} = (\boldsymbol{\eta}^{(0)}(\boldsymbol{D}), \forall \boldsymbol{D} \in \mathcal{D}) \text{ where } \boldsymbol{\eta}^{(0)}(\boldsymbol{D}) = (f(\mathrm{y}(v, \boldsymbol{D})), \forall v \in \mathcal{V}) \in \mathbb{R}^{|\mathcal{V}|}$$

Note that

$$\|\boldsymbol{\eta}^{(0)}\|_{\mathcal{H}}^2 = \sum_{v=1}^{|\mathcal{V}|} \int_{\boldsymbol{D}\in\mathcal{D}} |(f(\mathbf{y}(v,\boldsymbol{D}))|^2 d\mu(\boldsymbol{D}) \leq \hat{C}_1 \cdot \int_{\boldsymbol{D}\in\mathcal{D}} \|\boldsymbol{D}\|_2^2 d\mu(\boldsymbol{D}) < \infty$$

for some constant $\hat{C}_1 > 0$, where the first inequality follows from Assumption 5.4b and the fact that $f(\cdot)$ is a linear function, and the second inequality follows from the demand distribution has a bounded second moment. We conclude that $\eta_v^{(0)} \in L^2(\mathcal{D},\mu)$ for each $v = 1,\ldots,|\mathcal{V}|$ and thus $\boldsymbol{\eta}^{(0)} \in \mathcal{H}$. Moreover, since the function $f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{D}))$ is integrable over $\mathcal{D}\times\Phi$ with respect to the measure $\mu\times\lambda_k$, then we have

$$\begin{aligned}\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}))]d\lambda_k(\boldsymbol{\phi}) &= \int_{\boldsymbol{\phi}\in\Phi}\int_{\boldsymbol{D}\in\mathcal{D}} f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{D}))d\mu(\boldsymbol{D})d\lambda_k(\boldsymbol{\phi}) \\ &= \int_{\boldsymbol{D}\in\mathcal{D}}\int_{\boldsymbol{\phi}\in\Phi} f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\boldsymbol{D}))d\lambda_k(\boldsymbol{\phi})d\mu(\boldsymbol{D})\end{aligned}$$

where the second equality follows since Fubini's theorem implies that we can interchange the order of integral. Thus, it holds that

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}))]d\lambda_k(\boldsymbol{\phi}) = \int_{\boldsymbol{D}\in\mathcal{D}} \sum_{v=1}^{|\mathcal{V}|} \psi_v^k(D)\cdot\eta_v^{(0)}(\boldsymbol{D})d\mu(\boldsymbol{D}) = \langle\boldsymbol{\psi}^k,\boldsymbol{\eta}^{(0)}\rangle_{\mathcal{H}} \tag{D.17}$$

Thus, for each k, the expected allocation cost of the randomized policy λ_k can be expressed as the inner product of $\boldsymbol{\psi}^k$ and $\boldsymbol{\eta}^{(0)}$ in the space $\mathcal{H}$. Similarly, for each $j \in \mathcal{N}$, we define

$$\boldsymbol{\eta}^{(j)} = (\boldsymbol{\eta}^{(j)}(\boldsymbol{D}), \forall \boldsymbol{D}\in\mathcal{D}) \text{ where } \boldsymbol{\eta}^{(j)}(\boldsymbol{D}) = (R_j(s_j(v,\boldsymbol{D}),D_j), \forall v\in\mathcal{V}) \in \mathbb{R}^{|\mathcal{V}|}$$

Clearly, we have that

$$\|\boldsymbol{\eta}^{(j)}\|_{\mathcal{H}}^2 = \int_{\boldsymbol{D}\in\mathcal{D}} \|(R_j(s_j(v,\boldsymbol{D}),D_j),\forall v\in\mathcal{V})\|_2^2 d\mu(\boldsymbol{D}) \leq C_1\cdot\int_{\boldsymbol{D}\in\mathcal{D}} \|(\max\{1,D_j\},\forall j\in\mathcal{N})\|_2^2 d\mu(\boldsymbol{D}) < \infty$$

Then, we conclude that for each $j \in \mathcal{N}$, $\boldsymbol{\eta}^{(j)} \in \mathcal{H}$ and we have

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j)]d\lambda_k(\boldsymbol{\phi}) = \int_{D\in\mathcal{D}} \sum_{v=1}^{|\mathcal{V}|} \psi_v^k(D) \cdot \eta_v^{(j)}(\boldsymbol{D})d\mu(\boldsymbol{D}) = \langle \boldsymbol{\psi}^k, \boldsymbol{\eta}^{(j)} \rangle_{\mathcal{H}} \tag{D.18}$$

Thus, for each k, the service level for customer j obtained by the randomized policy λ_k can be expressed as the inner product of $\boldsymbol{\psi}^k$ and $\boldsymbol{\eta}^{(j)}$ in the space $\mathcal{H}$, for each $j \in \mathcal{N}$.

From (D.16), we know that the sequence $\{\boldsymbol{\psi}^k\}_{\forall k}$ is a bounded sequence in the Hilbert space $\mathcal{H}$. Then, from Banach-Saks theorem, there exists a subsequence $\{n_k\}_{\forall k}$ of $\{1, 2, \dots\}$ and $\hat{\boldsymbol{\psi}} \in \mathcal{H}$, such that the sequence $\{\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}\}_{\forall k}$ converges strongly to $\hat{\boldsymbol{\psi}}$ in the space $\mathcal{H}$. This implies that the sequence $\{\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}\}_{\forall k}$ converges weakly to $\hat{\boldsymbol{\psi}}$, then from (D.17), we have that

$$\begin{aligned} \langle \hat{\boldsymbol{\psi}}, \boldsymbol{\eta}^{(0)} \rangle_{\mathcal{H}} &= \lim_{k\to\infty} \langle \frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}, \boldsymbol{\eta}^{(0)} \rangle_{\mathcal{H}} = \lim_{k\to\infty} \frac{1}{k} \cdot \sum_{l=1}^{k} \langle \boldsymbol{\psi}^{n_l}, \boldsymbol{\eta}^{(0)} \rangle_{\mathcal{H}} \leq \limsup_{k\to\infty} \langle \boldsymbol{\psi}^k, \boldsymbol{\eta}^{(0)} \rangle_{\mathcal{H}} \\ &= \limsup_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))]d\lambda_k(\boldsymbol{\phi}) \end{aligned} \tag{D.19}$$

For each $j \in \mathcal{N}$, from (D.18), we have that

$$\langle \hat{\boldsymbol{\psi}}, \boldsymbol{\eta}^{(j)} \rangle_{\mathcal{H}} = \lim_{k\to\infty} \langle \frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}, \boldsymbol{\eta}^{(j)} \rangle_{\mathcal{H}} \geq \liminf_{k\to\infty} \langle \boldsymbol{\psi}^k, \boldsymbol{\eta}^{(j)} \rangle_{\mathcal{H}} = \liminf_{k\to\infty} \int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}), D_j)]d\lambda_k(\boldsymbol{\phi}) \tag{D.20}$$

It only remains to show that $\hat{\boldsymbol{\psi}}$ can be expressed as a randomized policy $\hat{\lambda}$. To that end, we first define a set containing all $\boldsymbol{D} \in \mathcal{D}$ such that $\hat{\boldsymbol{\psi}}(\boldsymbol{D})$ does not characterize a distribution over $\{a(v, \boldsymbol{D})\}_{\forall v\in\mathcal{V}}$:

$$\hat{\mathcal{D}} := \{\boldsymbol{D} \in \mathcal{D} : \hat{\boldsymbol{\psi}}(\boldsymbol{D}) \notin \mathcal{L}\}$$

where the set $\mathcal{L}$ is defined in (D.14). We have the following result.

Claim D.1.1. *It holds that* $\mu(\hat{\mathcal{D}}) = 0$.

For those $\boldsymbol{D} \in \hat{\mathcal{D}}$, we can simply change $\hat{\boldsymbol{\psi}}(\boldsymbol{D})$ into a point in the set $\mathcal{L}$. In this way, we

construct a $\hat{\boldsymbol{\psi}}$ such that $\hat{\boldsymbol{\psi}}(\boldsymbol{D}) \in \mathcal{L}$ for any $\boldsymbol{D} \in \mathcal{D}$. Since $\mu(\hat{\mathcal{D}}) = 0$, we conclude that (D.19) and (D.20) still hold.

Now we show that $\hat{\boldsymbol{\psi}}$ can be characterized as a randomized policy $\hat{\lambda}$. For each $\boldsymbol{D} \in \mathcal{D}$, we divide the interval $[0, 1]$ into a set of sub-intervals $\{I_v(\boldsymbol{D})\}_{v \in \mathcal{V}}$, such that $\hat{\psi}_v(\boldsymbol{D}) = |I_v(\boldsymbol{D})|$ for each $v \in \mathcal{V}$, where $|\cdot|$ denotes the length (Lebesgue measure) of the sub-interval. Note that for each $\boldsymbol{D} \in \mathcal{D}$, the randomized allocation specified by $\hat{\boldsymbol{\psi}}(\boldsymbol{D})$ can be interpreted as picking up a point x uniformly from the interval $[0, 1]$, and implementing the allocation $a(v, \boldsymbol{D})$ if and only if $x \in I_v(\boldsymbol{D})$. Thus, for each $x \in [0, 1]$, we can specify a deterministic policy

$$\boldsymbol{\phi}_x = (\boldsymbol{\phi}_x(\boldsymbol{D}), \forall \boldsymbol{D} \in \mathcal{D}) \text{ where } \boldsymbol{\phi}_x(\boldsymbol{D}) = a(v, \boldsymbol{D}) \text{ if and only if } x \in I_v(\boldsymbol{D})$$

We define the randomized policy $\hat{\lambda}$ as the uniform distribution over the set of deterministic policies $\{\boldsymbol{\phi}_x\}_{\forall x \in [0,1]}$. We then prove (D.6) and (D.7).

For the expected allocation cost, we have that

$$\begin{aligned}\int_{\boldsymbol{\phi} \in \Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi}, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\hat{\lambda}(\boldsymbol{\phi}) &= \int_{x \in [0,1]} \int_{\boldsymbol{D} \in \mathcal{D}} [f(\mathbf{y}(\boldsymbol{\phi}_x, \boldsymbol{c}, \tilde{\mathbf{D}}))] d\mu(\boldsymbol{D}) dx \\ &= \int_{\boldsymbol{D} \in \mathcal{D}} \int_{x \in [0,1]} [f(\mathbf{y}(\boldsymbol{\phi}_x, \boldsymbol{c}, \tilde{\mathbf{D}}))] dx \, d\mu(\boldsymbol{D})\end{aligned}$$

where the second equality follows by noting that $f(\mathbf{y}(\boldsymbol{\phi}_x, \boldsymbol{c}, \tilde{\mathbf{D}}))$ is integrable over $\mathcal{D} \times [0, 1]$, then Fubini's theorem implies that we can interchange the order of integration. Moreover, we have

$$\begin{aligned}\int_{\boldsymbol{D} \in \mathcal{D}} \int_{x \in [0,1]} [f(\mathbf{y}(\boldsymbol{\phi}_x, \boldsymbol{c}, \tilde{\mathbf{D}}))] dx \, d\mu(\boldsymbol{D}) &= \int_{\boldsymbol{D} \in \mathcal{D}} \sum_{v=1}^{|\mathcal{V}|} f(\mathbf{y}(v, \boldsymbol{D})) \cdot |I_v(D)| d\mu(\boldsymbol{D}) \\ &= \int_{\boldsymbol{D} \in \mathcal{D}} \sum_{v=1}^{|\mathcal{V}|} f(\mathbf{y}(v, \boldsymbol{D})) \cdot \hat{\psi}_v(\boldsymbol{D}) d\mu(\boldsymbol{D}) = \langle \hat{\boldsymbol{\psi}}, \boldsymbol{\eta}^{(0)} \rangle_{\mathcal{H}}\end{aligned}$$

Combing with (D.19), we have that

$$\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}))]d\hat{\lambda}(\boldsymbol{\phi}) = \langle\hat{\boldsymbol{\psi}},\boldsymbol{\eta}^{(0)}\rangle_{\mathcal{H}} \leq \limsup_{k\to\infty}\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[f(\mathbf{y}(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}))]d\lambda_k(\boldsymbol{\phi})$$

Similarly, for each $j \in \mathcal{N}$, we have that

$$\begin{aligned}
\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)]d\hat{\lambda}(\boldsymbol{\phi}) &= \int_{x\in[0,1]}\int_{\boldsymbol{D}\in\mathcal{D}}[R_j(s_j(\boldsymbol{\phi}_x,\boldsymbol{c},\tilde{\mathbf{D}}),D_j)]d\mu(\boldsymbol{D})dx \\
&= \int_{\boldsymbol{D}\in\mathcal{D}}\int_{x\in[0,1]}[R_j(s_j(\boldsymbol{\phi}_x,\boldsymbol{c},\tilde{\mathbf{D}}),D_j)]dx\, d\mu(\boldsymbol{D}) \\
&= \int_{\boldsymbol{D}\in\mathcal{D}}\sum_{v=1}^{|\mathcal{V}|} R_j(s_j(v,\boldsymbol{D}),D_j)\cdot|I_v(D)|d\mu(\boldsymbol{D}) = \langle\hat{\boldsymbol{\psi}},\boldsymbol{\eta}^{(j)}\rangle_{\mathcal{H}} \\
&\geq \liminf_{k\to\infty}\int_{\boldsymbol{\phi}\in\Phi} E_{\tilde{\mathbf{D}}}[R_j(s_j(\boldsymbol{\phi},\boldsymbol{c},\tilde{\mathbf{D}}),D_j)]d\lambda_k(\boldsymbol{\phi})
\end{aligned}$$

which completes our proof. □

Proof of Claim D.1.1:. We prove our result by contradiction. Suppose that $\mu(\hat{\mathcal{D}}) > 0$, then for each integer p, we define the set

$$\mathcal{D}_p = \{\boldsymbol{D}\in\mathcal{D} : \text{dist}(\hat{\boldsymbol{\psi}}(\boldsymbol{D}),\mathcal{L}) \geq \frac{1}{p}\}$$

where $\text{dist}(\hat{\boldsymbol{\psi}}(\boldsymbol{D}),\mathcal{L}) = \inf_{\mathbf{x}\in\mathcal{L}}\|\hat{\boldsymbol{\psi}}(\boldsymbol{D})-\mathbf{x}\|_2^2$ denoting the distance from the point $\hat{\boldsymbol{\psi}}(\boldsymbol{D}) \in \mathbb{R}^{|\mathcal{V}|}$ to the set $\mathcal{L} \subset \mathbb{R}^{|\mathcal{V}|}$. Since the set $\mathcal{L}$ is closed, for any $\boldsymbol{D} \in \hat{\mathcal{D}} = \{\boldsymbol{D}\in\mathcal{D} : \hat{\boldsymbol{\psi}}(\boldsymbol{D}) \notin \mathcal{L}\}$, it holds that $\text{dist}(\hat{\boldsymbol{\psi}}(\boldsymbol{D}),\mathcal{L}) > 0$, which implies that

$$\hat{\mathcal{D}} = \bigcup_{p=1}^{\infty}\mathcal{D}_p$$

Moreover, note that $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \cdots \subset \mathcal{D}_p \subset \ldots$, for each integer p, we define the set $\mathcal{E}_p = \mathcal{D}_p \setminus \mathcal{D}_{p-1} = \{\boldsymbol{D}\in\mathcal{D}_p : \boldsymbol{D}\notin\mathcal{D}_{p-1}\}$, where $\mathcal{D}_0 = \varnothing$. Obviously, the sets $\{\mathcal{E}_p\}$ are mutually

disjoint and it holds that

$$\hat{\mathcal{D}} = \bigcup_{p=1}^{\infty} \mathcal{E}_p \ \text{ and } \ \mathcal{D}_p = \bigcup_{l=1}^{p} \mathcal{E}_l \text{ for each integer } p$$

Then, from the coutable additivity of the measure μ, we have

$$\mu(\hat{\mathcal{D}}) = \mu(\bigcup_{p=1}^{\infty} \mathcal{E}_p) = \sum_{p=1}^{\infty} \mu(\mathcal{E}_p) = \lim_{p\to\infty} \sum_{l=1}^{p} \mu(\mathcal{E}_l) = \lim_{p\to\infty} \mu(\bigcup_{l=1}^{p} \mathcal{E}_l) = \lim_{p\to\infty} \mu(\mathcal{D}_p)$$

Thus, we conclude that there exists an integer p_0, such that $\mu(\mathcal{D}_{p_0}) > 0$.

On the other hand, the sequence $\{\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}\}_{\forall k}$ converges strongly to $\hat{\boldsymbol{\psi}}$ implies that the sequence $\{\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}\}_{\forall k}$ converges to $\hat{\boldsymbol{\psi}}$ in measure, i.e., for any integer p,

$$\lim_{k\to\infty} \mu\left(\left\{\boldsymbol{D} \in \mathcal{D} : \|\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}(\boldsymbol{D}) - \hat{\boldsymbol{\psi}}(\boldsymbol{D})\|_2^2 \geq \frac{1}{p}\right\}\right) = 0$$

Moreover, note that $\mathcal{L}$ is a convex set, then for each $\boldsymbol{D} \in \mathcal{D}$, we must have $\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}(\boldsymbol{D}) \in \mathcal{L}$ for each k. Thus, it holds that

$$\text{dist}(\hat{\boldsymbol{\psi}}(\boldsymbol{D}), \mathcal{L}) \leq \|\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}(\boldsymbol{D}) - \hat{\boldsymbol{\psi}}(\boldsymbol{D})\|_2^2, \ \ \forall k$$

which implies that the set $\mathcal{D}_p$ is contained in the set $\left\{\boldsymbol{D} \in \mathcal{D} : \|\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}(\boldsymbol{D}) - \hat{\boldsymbol{\psi}}(\boldsymbol{D})\|_2^2 \geq \frac{1}{p}\right\}$ for each k. As a result, we have that

$$\mu(\mathcal{D}_p) \leq \mu\left(\left\{\boldsymbol{D} \in \mathcal{D} : \|\frac{1}{k} \cdot \sum_{l=1}^{k} \boldsymbol{\psi}^{n_l}(\boldsymbol{D}) - \hat{\boldsymbol{\psi}}(\boldsymbol{D})\|_2^2 \geq \frac{1}{p}\right\}\right), \ \ \forall k$$

and thus $\mu(\mathcal{D}_p) = 0$ for any integer p, which is a contradiction. □

Proof of Theorem 5.7

We first present the two well-known results, which will be useful for our proof of Theorem 5.7.

Lemma D.1.1. *(Azuma's Inequality (Azuma, 1967)) Suppose* $\{X_k, k = 0, 1, 2, \dots\}$ *is a martingale and* $|X_k - X_{k-1}| < c_k$ *almost surely for each* k*, then for all positive integer* N *and all positive real* ϵ*,*

$$P(|X_N - X_0| > \epsilon) \leq 2 \cdot \exp\left(\frac{-\epsilon^2}{2\sum_{k=1}^{N} c_k^2}\right)$$

Lemma D.1.2. *(Borel-Cantelli Lemma (Borel, 1909)) Let* $E_1, E_2, \dots$ *be a sequence of events in a probability space. If*

$$\sum_{n=1}^{\infty} P(E_n) < \infty$$

then

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k\right) = 0$$

where $\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k$ *denotes the set of outcomes that occur infinite times with the sequence* $\{E_k\}_{k\geq 1}$*.*

Now we are ready to prove Theorem 5.7. The main idea of the proof is to show that by following the dual update step of the Max-Weighted-Service policy, the gap between the expected cost of our policy and the optimal value of (5.11), as well as the gap between the achieved and the target service level, can both be bounded by some functions of the dual variables, which diminish under carefully chosen step sizes, as $T \to \infty$.

Proof of Theorem 5.7:. We first prove (5.21). From weak duality (Shapiro, 2001), it holds that $G(\mathbf{w}^*) \leq$ Obj (5.11). Thus, it is enough to prove that

$$\limsup_{T\to\infty} \frac{1}{T} \cdot \mathrm{E}_{\tilde{\mathbf{D}}^t}[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] \leq G(\mathbf{w}^*)$$

holds almost surely. Note that by definition of $G(\cdot)$ and $\mathbf{w}^*$, we have that

$$G(\mathbf{w}^*) = \max_{\mathbf{w}\geq 0} G(\mathbf{w}) \geq \frac{1}{T}\cdot\sum_{t=1}^{T} G(\mathbf{w}^{(t)})$$
$$\geq \frac{1}{T}\cdot\sum_{t=1}^{T}\left(\sum_{j\in\mathcal{N}} w_j^{(t)}\cdot\beta_j + \mathrm{E}_{\tilde{\mathbf{D}}^t}[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] - \sum_{j\in\mathcal{N}} w_j^{(t)}\cdot\ \mathrm{E}_{\tilde{\mathbf{D}}^t}[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t)]\right)$$

Re-arranging terms, we have

$$\frac{1}{T}\cdot\sum_{t=1}^{T}\mathrm{E}_{\tilde{\mathbf{D}}^t}[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] - G(\mathbf{w}^*) \leq \frac{1}{T}\cdot\sum_{t=1}^{T}\left(-\sum_{j\in\mathcal{N}} w_j^{(t)}\cdot\beta_j + \sum_{j\in\mathcal{N}} w_j^{(t)}\cdot\ \mathrm{E}_{\tilde{\mathbf{D}}^t}[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t)]\right) \tag{D.21}$$

We now proceed to upper bound the right hand side of the above inequality. From the update rule (5.20), we have that

$$\begin{aligned}
\|\mathbf{w}^{(t+1)}\|^2 &\leq \left\|\left(w_j^{(t)} + \gamma_T\cdot\left(\beta_j - R_j\left(s_j\left(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}\right), D_j^{(t)}\right)\right),\ j\in\mathcal{N}\right)\right\|^2 \\
&= \|\mathbf{w}^{(t)}\|^2 + \gamma_T^2\cdot\left\|\left(\beta_j - R_j\left(s_j\left(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}\right), D_j^{(t)}\right),\ j\in\mathcal{N}\right)\right\|^2 \\
&+ 2\gamma_T\cdot\sum_{j\in\mathcal{N}}\left(w_j^{(t)}\cdot\ \mathrm{E}_{\boldsymbol{D}^{(t)}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right] - w_j^{(t)}\cdot R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right) \\
&+ 2\gamma_T\cdot\sum_{j\in\mathcal{N}}\left(w_j^{(t)}\cdot\beta_j - w_j^{(t)}\cdot\ \mathrm{E}_{\boldsymbol{D}^{(t)}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right]\right)
\end{aligned} \tag{D.22}$$

where $\|\cdot\|$ denotes the L_2 norm and we denote $\left(a_j,\ j\in\mathcal{N}\right)$ as a n dimensional vector with a_j on its j-th component for any $j\in\mathcal{N}$. Moreover, for each t, we denote

$$L_t = \frac{2}{T\gamma_T}\cdot\sum_{j\in\mathcal{N}}\left(w_j^{(t)}\cdot\ \mathrm{E}_{\boldsymbol{D}^{(t)}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right] - w_j^{(t)}\cdot R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right) \tag{D.23}$$

Obviously, $\{L_t\}_{t\geq 1}$ is a sequence of martingale difference with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 1}$, where for each t, $\mathcal{F}_t$ denotes the σ-algebra $\sigma\left(\boldsymbol{D}^{(1)}, \ldots, \boldsymbol{D}^{(t)}\right)$. From the update rule (5.20), we have $w_j^{(t+1)} \leq t\cdot\gamma_T\cdot\beta_j$ for each $j\in\mathcal{N}$ and each t. Thus, by Assumption 5.6, there exists a constant

$\hat{C}_1$ such that $|L_t| \leq \hat{C}_1$ almost surely for each t. Note that it follows from Assumption 5.6 that for any $\mathbf{w}^{(t)}$ and $\tilde{\mathbf{D}}$, we have

$$\| \left(\beta_j - R_j \left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}), \tilde{D}_j \right), \ j \in \mathcal{N} \right) \|_2^2 \leqslant n(1+C)^2. \tag{D.24}$$

Further note that for each t, we have

$$w_j^{(t)} \cdot \mathrm{E}_{\boldsymbol{D}^{(t)}} \left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)}) \right] = w_j^{(t)} \cdot \mathrm{E}_{\tilde{\mathbf{D}}^t} \left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t) \right] \quad (\text{since } \tilde{\mathbf{D}}^t \sim \boldsymbol{D}^{(t)})$$

Then, (D.22) implies that

$$\|\mathbf{w}^{(t+1)}\|^2 \leq \|\mathbf{w}^{(t)}\|^2 + \gamma_T^2 \cdot n(1+C)^2 + T\gamma_T^2 \cdot L_t + 2\gamma_T \cdot \sum_{j \in \mathcal{N}} \left(w_j^{(t)} \cdot \beta_j - w_j^{(t)} \cdot \mathrm{E}_{\tilde{\mathbf{D}}^t} \left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{\mathbf{D}}_j^t) \right] \right)$$

and thus, by re-arranging terms, we have

$$\sum_{j \in \mathcal{N}} \left(-w_j^{(t)} \cdot \beta_j + w_j^{(t)} \cdot \mathrm{E}_{\tilde{\mathbf{D}}^t} \left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{\mathbf{D}}_j^t) \right] \right) \geqslant \frac{1}{2\gamma_T} \left(-\|\mathbf{w}^{(t+1)}\|^2 + \|\mathbf{w}^{(t)}\|^2 + \gamma_T^2 \cdot n(1+C)^2 + T\gamma_T^2 \cdot L_t \right) \tag{D.25}$$

Plugging (D.25) into (D.21), we have that

$$\begin{aligned}
&\frac{1}{T} \cdot \sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathbf{D}}^t} [f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t))] - G(\mathbf{w}^*) \\
\leq &\frac{1}{T} \cdot \sum_{t=1}^{T} \left(- \sum_{j \in \mathcal{N}} w_j^{(t)} \cdot \beta_j + \sum_{j \in \mathcal{N}} w_j^{(t)} \cdot \mathrm{E}_{\tilde{\mathbf{D}}^t} [R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t)] \right) \\
\leq &\frac{1}{2T\gamma_T} \cdot \sum_{t=1}^{T} \left\{ \|\mathbf{w}^{(t)}\|^2 - \|\mathbf{w}^{(t+1)}\|^2 + \gamma_T^2 \cdot n(1+C)^2 + T\gamma_T^2 \cdot L_t \right\} \\
= &\frac{1}{2T\gamma_T} \cdot (\|\mathbf{w}^{(1)}\|^2 - \|\mathbf{w}^{(T+1)}\|^2) + \frac{\gamma_T \cdot n(1+C)^2}{2} + \frac{\gamma_T}{2} \cdot \sum_{t=1}^{T} L_t \\
\leq &\frac{\gamma_T \cdot n(1+C)^2}{2} + \frac{\gamma_T}{2} \cdot \sum_{t=1}^{T} L_t
\end{aligned} \tag{D.26}$$

where the last inequality holds since $\mathbf{w}^{(1)} = 0$.

For any $a > 0$, define $E_T(a)$ as the event that $\gamma_T \cdot |\sum_{t=1}^{T} L_t| \geq a$. Since we have shown $|L_t| \leq \hat{C}_1$ almost surely for each t and $\gamma_T = T^{-(\frac{1}{2}+\epsilon)}$ for some $\epsilon \in (0, 1/2)$, by Azuma's inequality, we have

$$P(E_T(a)) = P(\gamma_T \cdot |\sum_{t=1}^{T} L_t| \geq a) \leq 2 \cdot \exp(-\frac{a^2}{2\hat{C}_1^2} \cdot T^{2\epsilon})$$

Note that the above inequality implies that

$$\sum_{T=1}^{\infty} P(E_T(a)) \leq 2 \cdot \sum_{T=1}^{\infty} \exp(-\frac{a^2}{2\hat{C}_1^2} \cdot T^{2\epsilon}) < \infty$$

Then, by Borel-Cantelli Lemma, we know that $P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k(a)\right) = 0$ for each $a > 0$. Thus, it holds that

$$\lim_{T\to\infty} \gamma_T \cdot \sum_{t=1}^{T} L_t = 0 \quad \text{almost surely}$$

when $\gamma_T = T^{-(\frac{1}{2}+\epsilon)}$ for some $\epsilon \in (0, 1/2)$, which completes our proof of (5.21).

We then prove (5.22). To that end, we define, for each $j \in \mathcal{N}$ and each $\tau = 1, \ldots, T$,

$$\rho_{j,\tau} \quad = \quad \beta_j - \frac{1}{\tau} \sum_{t=1}^{\tau} R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^t), \tilde{D}_j^t).$$

Then (5.22) follows if $\limsup_{T\to\infty} \rho_{j,T} \leqslant 0$ almost surely for all $j \in \mathcal{N}$. To that end, we first prove that $\tau\rho_{j,\tau} \preceq \frac{w_j^{(\tau+1)}}{\gamma_T}$, $\forall j \in \mathcal{N}$ for any $\tau \geqslant 1$, where $a \preceq b$ denotes random variable a is first-order stochastic dominated by random variable b. Then, in Lemma D.1.3 below, we show that for each $j \in \mathcal{N}$, $\limsup_{T\to\infty} \frac{w_j^{(T+1)}}{T\gamma_T} = 0$ almost surely.

We prove $\tau\rho_{j,\tau} \preceq \frac{1}{\gamma_T} \cdot w_j^{(\tau+1)}$ by induction. When $\tau = 1$, noticing that $\mathbf{w}^{(1)} = 0$, we have that

$$\begin{aligned}
\rho_{j,1} &= \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(1)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^1), \tilde{D}_j^1\right) \\
&= w_j^{(1)} + \beta_j - \left[R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(1)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^1), \tilde{D}_j^1\right)\right] \\
&\leq \left[w_j^{(1)} + \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(1)}}, \boldsymbol{c}, \boldsymbol{d}^{(1)}), D_j^{(1)}\right)\right]^+ \quad (\text{since } \boldsymbol{D}^{(1)} \sim \tilde{\mathbf{D}}^1) \\
&= \frac{1}{\gamma_T} \cdot w_j^{(2)}
\end{aligned}$$

Now assume that we have $(\tau-1)\rho_{j,\tau-1} \preceq \frac{1}{\gamma_T} \cdot w_j^{(\tau)}$, then

$$\begin{aligned}
\tau\rho_{j,\tau} &= (\tau-1)\rho_{j,\tau-1} + \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(\tau)}}, \boldsymbol{c}, \tilde{\mathbf{D}}^\tau), \tilde{D}_j^\tau\right) \\
&\leq (\tau-1)\rho_{j,\tau-1} + \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(\tau)}}, \boldsymbol{c}, \boldsymbol{D}^{(\tau)}), \tilde{D}_j^{(\tau)}\right) \quad (\text{since } \boldsymbol{D}^{(\tau)} \sim \tilde{\mathbf{D}}^\tau) \\
&\leq \frac{1}{\gamma_T} \cdot w_j^{(\tau)} + \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(\tau)}}, \boldsymbol{c}, \boldsymbol{D}^{(\tau)}), \tilde{D}_j^{(\tau)}\right) \\
&\leq \left[\frac{1}{\gamma_T} \cdot w_j^{(\tau)} + \beta_j - R_j\left(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(\tau)}}, \boldsymbol{c}, \boldsymbol{D}^{(\tau)}), \tilde{D}_j^{(\tau)}\right)\right]^+ \\
&= \frac{1}{\gamma_T} \cdot w_j^{(\tau+1)}
\end{aligned}$$

Thus $T\rho_{j,T} \preceq \frac{1}{\gamma_T} \cdot w_j^{(T+1)}$, $\forall j \in \mathcal{N}$, which completes the proof. □

Lemma D.1.3. *If $\boldsymbol{c}$ is feasible, then for any $j \in \mathcal{N}$, $\limsup_{T\to\infty} \frac{w_j^{(T+1)}}{T\gamma_T} = 0$ almost surely.*

Proof:. It follows from (D.22) and Assumption 5.6 that

$$\begin{aligned}
\|\mathbf{w}^{(t+1)}\|^2 \leq \|\mathbf{w}^{(t)}\|^2 &+ \gamma_T^2 \cdot n(1+C)^2 + T\gamma_T^2 \cdot L_t \\
&+ 2\gamma_T \cdot \sum_{j\in\mathcal{N}} \left(w_j^{(t)} \cdot \beta_j - w_j^{(t)} \cdot \mathrm{E}_{\boldsymbol{D}^{(t)}}\left[R(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}}, \boldsymbol{c}, \boldsymbol{D}^{(t)}), D_j^{(t)})\right]\right)
\end{aligned}$$

Notice that

$$\mathrm{E}_{\boldsymbol{D}^{(t)}}\left[\sum_{j\in\mathcal{N}} w_j^{(t)}\cdot R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\boldsymbol{D}^{(t)}),D_j^{(t)})\right]=\sum_{j\in\mathcal{N}} w_j^{(t)}\cdot\ \mathrm{E}_{\tilde{\mathbf{D}}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)\right]\quad(\boldsymbol{D}^{(t)}\sim\tilde{\mathbf{D}})$$

Also, from weak duality, it holds $G(\mathbf{w}^{(t)})\le G(\mathbf{w}^*)\le$ Obj (5.11). Then, we have that

$$\sum_{j\in\mathcal{N}} w_j^{(t)}\beta_j+\ \mathrm{E}_{\tilde{\mathbf{D}}}\left[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\tilde{\mathbf{D}}))\right]-\sum_{j\in\mathcal{N}} w_j^{(t)}\ \mathrm{E}_{\tilde{\mathbf{D}}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)\right]\le\text{Obj (5.11)}$$

When the capacity level $\boldsymbol{c}$ is feasible, the objective value of (5.11) is finite and we denote $\hat{C}_2$ as its upper bound. Thus, we have that

$$\sum_{j\in\mathcal{N}} w_j^{(t)}\beta_j-\sum_{j\in\mathcal{N}} w_j^{(t)}\ \mathrm{E}_{\tilde{\mathbf{D}}}\left[R_j(s_j(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\tilde{\mathbf{D}}),\tilde{D}_j)\right]\le\text{Obj (5.11)}-\ \mathrm{E}_{\tilde{\mathbf{D}}}\left[f(\mathbf{y}(\boldsymbol{\phi}_{\mathbf{w}^{(t)}},\boldsymbol{c},\tilde{\mathbf{D}}))\right]\leqslant\hat{C}_2 \tag{D.27}$$

Therefore, we must have

$$\|\mathbf{w}^{(t+1)}\|_2^2\le\|\mathbf{w}^{(t)}\|_2^2+\gamma_T^2\cdot n(1+C)^2+2\gamma_T\cdot\hat{C}_2+T\gamma_T^2\cdot L_t\quad\forall t=1,2,\ldots,T \tag{D.28}$$

Summing inequality (D.28) from $t=1$ to T, we get

$$\|\mathbf{w}^{(T+1)}\|_2^2\le\|\mathbf{w}^{(1)}\|_2^2+T\gamma_T^2\cdot n(1+C)^2+2T\gamma_T\cdot\hat{C}+T\gamma_T^2\cdot\sum_{t=1}^{T}L_t$$

Then, we have

$$\frac{1}{T^2\gamma_T^2}\cdot\|\mathbf{w}^{(T+1)}\|_2^2\le\frac{1}{T}\cdot n(1+C)^2+\frac{2\hat{C}}{T\gamma_T}+\frac{1}{T}\cdot\sum_{t=1}^{T}L_t \tag{D.29}$$

Since we have shown $|L_t| \leq \hat{C}_1$ almost surely for each t, we can again apply the combination of Azuma's inequality and the Borel-Cantelli lemma to show that

$$\lim_{T\to\infty} \frac{1}{T} \cdot \sum_{t=1}^{T} L_t = 0 \quad \text{almost surely}$$

Thus, we have for each $j \in \mathcal{N}$,

$$\limsup_{T\to\infty} \frac{w_j^{(T+1)}}{T\gamma_T} = 0 \quad \text{almost surely}$$

which completes the proof. □

Proof of Corollary 5.8.. The proof of Corollary 5.8 follows the same main steps as that of Theorem 5.7 with minor modifications outlined below. Notice that Corollary 5.8, which focuses only on the expected performance of Algorithm 1, is weaker than Theorem 5.7. Thus, Assumption 3, which is needed for the proof of Theorem 5.7, is now replaced with a weaker condition, i.e., $E_D[R_j(s_j, \tilde{D}_j)^2] \leq C$ for all $s_j \geq 0$.

In particular, Assumption 3 is only used in Theorem 5.7 to prove the boundedness of L_t and inequality (D.24). For the proof of Corollary 5.8, we can take expectation over $w^{(t)}$ and $D^{(t)}$ on both sides of (D.22) and (D.24). Then (D.22) implies that $E_{w^{(t)},D^{(t)}}[L_t] = 0$ and thus bounded for each t . Also, inequality (D.24) holds in expectation as long as $E_D[R_j(s_j, \tilde{D}_j)^2] \leq C$ for all $s_j \geq 0$, which is the condition assumed in Corollary 5.8.

Then, (D.26) together with $\mathrm{E}_{\mathbf{w}^{(t)},\tilde{\mathbf{D}}^{(t)}}[L_t] = 0$, directly imply the convergence rate on the expected allocation cost should be $O(\gamma_T)$.

Finally, (D.29) implies, for any j,

$$\mathrm{E}[\frac{w_j^{(T+1)}}{T\gamma_T}] \leqslant O(\max\{\sqrt{\frac{1}{T}}, \sqrt{\frac{1}{T\gamma_T}}\}).$$

Thus, we have

$$\mathrm{E}[\rho_{j,T}] \leqslant \mathrm{E}[\frac{w_j^{(T+1)}}{T\gamma_T}] \leqslant O(\max\{\sqrt{\frac{1}{T}}, \sqrt{\frac{1}{T\gamma_T}}\}),$$

which proves the convergence rate on the expected service level. □

Proof of Theorem 5.9:. We denote by $\mathrm{Obj}_{\boldsymbol{c}}(5.11)$ the objective value of (5.11) given capacity level $\boldsymbol{c}$. From Theorem 5.5, we have that

$$\max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c}) = \begin{cases} p(\boldsymbol{c}) + \mathrm{Obj}_{\boldsymbol{c}}(5.11), & \text{if (5.11) is feasible for } \boldsymbol{c} \\ +\infty, & \text{if (5.11) is infeasible for } \boldsymbol{c} \end{cases}$$

Denote by $\boldsymbol{c}^*$ one optimal solution of (5.10). Then, we have

$$\mathrm{Obj}\ (5.10) = p(\boldsymbol{c}^*) + \mathrm{Obj}_{\boldsymbol{c}^*}(5.11) = \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c}^*) \geq \min_{\boldsymbol{c}\geq 0} \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c}) \tag{D.30}$$

Denote by $\hat{\boldsymbol{c}}$ one optimal solution of (5.24). Then, we have that

$$\min_{\boldsymbol{c}\geq 0} \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c}) = \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \hat{\boldsymbol{c}}) < +\infty$$

which implies that (5.11) is feasible under the capacity level $\hat{c}$. Thus, we have

$$\min_{\boldsymbol{c}\geq 0} \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c}) = \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \hat{\boldsymbol{c}}) = p(\hat{\boldsymbol{c}}) + \mathrm{Obj}_{\hat{\boldsymbol{c}}}(5.11) \geq \mathrm{Obj}\ (5.10) \tag{D.31}$$

As a result, we have $\mathrm{Obj}\ (5.10) = \min_{\boldsymbol{c}\geq 0} \max_{\mathbf{w}\geq 0} H(\mathbf{w}, \boldsymbol{c})$, and all the inequalities in (D.30) and (D.31) hold as equality, which implies that $\boldsymbol{c}^*$ is an optimal solution to (5.24) and $\hat{\boldsymbol{c}}$ is an optimal solution to (5.10). □

Proof of Lemma 5.9.1.. By definition, for any $\mathbf{w} \geqslant 0$ and $\boldsymbol{d}$,

$$g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{d}) = \begin{array}{ll} \min & f(\mathbf{y}) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j, D_j) \\ \text{s.t.} & (\boldsymbol{s}, \mathbf{y}) \in P(\boldsymbol{c}, \boldsymbol{D}) \end{array} \tag{D.32}$$

Let $(s^*(\boldsymbol{c}), y^*(\boldsymbol{c}))$ be an optimal solution. (Here we assume that $\mathbf{w}$ and $\boldsymbol{d}$ are fixed and thus drop the dependence on them in $(s^*(\boldsymbol{c}), y^*(\boldsymbol{c}))$.) Then we have $g(\mathbf{w}, \boldsymbol{c}; \boldsymbol{d}) = f(\mathbf{y}) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j^*(\boldsymbol{c}), \boldsymbol{d})$. For any two points $\boldsymbol{c}^1, \boldsymbol{c}^2$ and any constant $0 < \alpha < 1$,

$$(\alpha s^*(\boldsymbol{c}^1) + (1-\alpha) s^*(\boldsymbol{c}^2)), ((\alpha y^*(\boldsymbol{c}^1) + (1-\alpha) y^* \boldsymbol{c}^2)))$$

must be a feasible solution to (D.32) when $\boldsymbol{c} = \alpha \boldsymbol{c}^1 + (1-\alpha)\boldsymbol{c}^2$. Then we have $(\alpha s^*(\boldsymbol{c}^1) + (1-\alpha) s^*(\boldsymbol{c}^2), \alpha y^*(\boldsymbol{c}^1) + (1-\alpha) y^*(\boldsymbol{c}^2)) \in P((\alpha \boldsymbol{c}^1 + (1-\alpha)\boldsymbol{c}^2), \boldsymbol{d})$. Thus, from the concavity of $R_j(s_j, D_j)$ in s_j, we have that

$$\begin{aligned}
& \alpha g(\mathbf{w}, \boldsymbol{c}^1; \boldsymbol{d}) + (1-\alpha) g(\mathbf{w}, \boldsymbol{c}^2; \boldsymbol{d}) \\
= \; & (\alpha \cdot f(\mathbf{y}^*(\boldsymbol{c}^1)) + (1-\alpha) \cdot f(\mathbf{y}^*(\boldsymbol{c}^2))) - \sum_{j \in \mathcal{N}} w_j \cdot (\alpha R_j(s_j^*(\boldsymbol{c}^1), D_j) + (1-\alpha) R_j(s_j^*(\boldsymbol{c}^2), D_j)) \\
\geq \; & f(\alpha \cdot \mathbf{y}^*(\boldsymbol{c}^1) + (1-\alpha) \cdot \mathbf{y}^*(\boldsymbol{c}^2)) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(\alpha s_j^*(\boldsymbol{c}^1) + (1-\alpha) s_j^*(\boldsymbol{c}^2), D_j) \\
\geq \; & \min_{(\boldsymbol{s}, \mathbf{y}) \in P(\alpha \boldsymbol{c}^1 + (1-\alpha)\boldsymbol{c}^2, \boldsymbol{d})} f(\mathbf{y}) - \sum_{j \in \mathcal{N}} w_j \cdot R_j(s_j, D_j) \\
= \; & g(\mathbf{w}, \alpha \boldsymbol{c}^1 + (1-\alpha)\boldsymbol{c}^2; \boldsymbol{d})
\end{aligned}$$

We conclude that $g(\mathbf{w}, \boldsymbol{c}; \tilde{\mathbf{D}})$ is convex in $\boldsymbol{c}$ for any $\mathbf{w} \geqslant 0$ and $\tilde{\mathbf{D}}$, and thus $H(\mathbf{w}, \boldsymbol{c})$ is a convex function of $\boldsymbol{c}$ for any $\mathbf{w} \geqslant 0$. □

E | Appendix for Chapter 6

Useful Previous Results

We first state the well-known Hoeffdeing's inequality, which establishes concentration bound for i.i.d. random variables.

Lemma E.0.1 (Hoeffding's Inequality). *Let $X_1, \ldots, X_m$ be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely for each $i \in [m]$. Then, denote by $S_n = \sum_{i=1}^m X_i$. It holds that*

$$P(S_n - \mathbb{E}[S_n] \geq t) \leq \exp(-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2})$$

We then state a general concentration bound for Markov chain with stationary distributions from Healy (2008).

Lemma E.0.2 (Theorem 1.1 in Healy (2008)). *Let $\boldsymbol{X} = (X_i, i \geq 1)$ be a Markov chain with a stationary distribution ϕ. Suppose that the distribution of X_1 is identical to the distribution of ϕ. Then, there exists a constant $\lambda > 0$ such that for any $\epsilon > 0$, it holds that*

$$P\left(\left|\sum_{i=1}^m X_i - \mathbb{E}\left[\sum_{i=1}^m X_i\right]\right| \geq \sqrt{m} \cdot \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \tag{E.1}$$

We also state the following lemma regarding the convexity of the pseudo-cost.

Lemma E.0.3 (Proposition 1 in Bu et al. (2020)). *We denote by* $\hat{s}(\mu, Z) = s(q(\mu), z)$ *where* $q(\mu) = \min_q\{q : \mathbb{E}[s(q, Z)] \geq \mu\}$. *We also denote the transformed cost function* $TC(\mu) = \hat{C}_\infty^{\pi_{q(\mu)}}$. *Suppose that the random supply function takes one of the four formulations specified in* **??**. *Then,* $TC(\mu)$ *is a convex function over* $[0, \bar{\mu}]$, *where* $\bar{\mu}$ *satisfying* $q(\bar{\mu}) = \bar{q}$.

We finally state the following result, showing how the limiting inventory level can be bounded.

Lemma E.0.4 (Lundberg's Inequality). *Denote by* I_∞ *as the limiting distribution of the stochastic process* $I_{t+1} = (I_t + Q - D)^+$, *where* Q *and* D *are two positive random variables. Then, there exists a constant* ρ *such that for any* $a > 0$, *we have*

$$P(I_\infty \geq a) \leq \exp(-\rho a).$$

Moreover, ρ *is the adjustment coefficient of the random variable* $Q - D$, *which is defined as the solution to* $\lambda(z) = 1$, *where* $\lambda(z) = \mathbb{E}[\exp(z \cdot (Q - D))]$.

Missing Proofs

Proof of Lemma 6.3.2. From Lemma E.0.3, we know that the transformed cost function $\mathrm{TC}(\mu) = \hat{C}_\infty^{\pi_{q(\mu)}}$ is a convex function over $\mu \in [0, \bar{\mu}]$, which is a bounded region. Thus, we know that there exists a constant $\beta' > 0$ such that

$$|\mathrm{TC}(\mu_1) - \mathrm{TC}(\mu_2)| \leq \beta' \cdot |\mu_1 - \mu_2|, \ \forall \mu_1, \mu_2 \in [0, \bar{\mu}]. \tag{E.2}$$

For any $q_1, q_2 \in [0, \bar{q}]$, we now denote by $\mu_1 = \mathbb{E}[s(q_1, Z)]$ and $\mu_2 = \mathbb{E}[s(q_2, Z)]$. Moreover, for the random supply function taking one of the four formulations specified in **??**, it is direct to check that there exists a constant $\alpha' > 0$ such that

$$|\mu_1 - \mu_2| \leq \alpha' \cdot |q_1 - q_2| \tag{E.3}$$

Plugging (E.3) into (E.2), we know that

$$|\hat{C}_{\infty}^{\pi_{q_1}} - \hat{C}_{\infty}^{\pi_{q_2}}| = |\mathrm{TC}(\mu_1) - \mathrm{TC}(\mu_2)| \leq \beta' \cdot |\mu_1 - \mu_2| \leq \alpha'\beta' \cdot |q_1 - q_2|$$

Therefore, we prove that $\hat{C}_{\infty}^{\pi_q}$ is Lipschitz continuous over q with a Lipschitz constant $\beta = \alpha'\beta'$.

□

Proof of Lemma 6.3.3. For each epoch n, we denote by

$$\mathcal{B}_n = \{I_{\tau_{n'}} \leq \kappa_1 \cdot \log T,\ \tilde{I}^{a^{n*}}_{\tau_{n'}} \leq \kappa_1 \cdot \log T \text{ and } I_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{n'*}}_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}},\ \forall n' \leq n\}.$$

In order to prove the lemma, it is sufficient to prove that

$$P(\mathcal{B}_n) \geq 1 - \frac{3n}{T^2}. \tag{E.4}$$

We prove (E.4) by using induction on the epoch n.

Clearly, when $n = 1$, we have that $P(I_{\tau_1} = 0) = 1$. From Lemma E.0.4, there exists a constant $\kappa_1 > 0$ such that

$$P(\tilde{I}^{a^{1*}}_{\tau_1} \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2}$$

by noting that the distribution of $\tilde{I}^{a^{1*}}_{\tau_1}$ is identical to the distribution of $I^{a^{1*}}_{\infty}$.

Now conditioning on the event $\{\tilde{I}^{a^{1*}}_{\tau_1} \leq \kappa_1 \cdot \log T\}$, we proceed to bound the probability that event $\{I_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{1*}}_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}}\}$ happens. Note that the evolution of the stochastic process I_t in (6.10) is identical to the evolution of the stochastic process $\tilde{I}^{a^{1*}}_t$ in (6.13) for $t = \tau_1, \ldots, \tau_2 - 1$. Therefore, it is clear to see that the event $\{I_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{1*}}_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}}\}$ happens as long as

$$I_t = \tilde{I}^{a^{1*}}_t = 0, \quad \text{for some } t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}. \tag{E.5}$$

We note that $I_{\tau_1} \leq \tilde{I}_{\tau_1}^{a^{1*}}$ implies that $I_t \leq \tilde{I}_t^{a^{1*}}$ for all $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. From the non-negativity of I_t and $\tilde{I}_t^{a^{1*}}$, we have that (E.5) holds as long as $\tilde{I}_t^{a^{1*}} = 0$ for some $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. As a result, a sufficient condition for (E.5) to hold is that

$$\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a^{1*}, Z_t) \geq \kappa_1 \cdot \log T$$

Note that $D_t - s(a^{1*}, Z_t)$ are i.i.d. random varibles for $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. We also denote by $\delta = \mathbb{E}[D] - \mathbb{E}[s(\bar{q}, Z)]$. Following Hoeffding's inequality (Lemma E.0.1), we have that

$$P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a^{1*}, Z_t) \geq \kappa_1 \cdot \log T\right) \geq 1 - \exp\left(-\frac{2(\delta\kappa_2 \max\{\log T, 2L\} - \kappa_1 \log T)^2}{\kappa_2 \cdot \max\{\log T, 2L\} \cdot \bar{D}}\right) \geq 1 - \frac{1}{T^2}$$

where $\kappa_2 \geq \max\{\frac{2\kappa_1}{\delta}, \frac{4\bar{D}}{\delta^2}\} \geq \max\{\frac{2\kappa_1 \log T}{\delta \cdot \max\{\log T, 2L\}}, \frac{4\bar{D}\log T}{\delta^2 \cdot \max\{\log T, 2L\}}\}$.

The above derivation implies that

$$\begin{aligned}
P(\mathcal{B}_1) &= P\left(\mathcal{B}_1 \mid \tilde{I}_{\tau_1}^{a^{1*}} \leq \kappa_1 \cdot \log T\right) \cdot P(\tilde{I}_{\tau_1}^{a^{1*}} \leq \kappa_1 \cdot \log T) \\
&= P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a^{1*}, Z_t) \geq \kappa_1 \cdot \log T\right) \cdot P(\tilde{I}_{\tau_1}^{a^{1*}} \leq \kappa_1 \cdot \log T) \\
&\geq (1 - \frac{1}{T^2}) \cdot (1 - \frac{1}{T^2}) \geq 1 - \frac{2}{T^2} \geq 1 - \frac{3}{T^2}
\end{aligned}$$

Therefore, we prove (E.4) for $n = 1$.

Now assume that (E.4) holds for epoch $n-1$. We consider epoch n. Clearly, from the definition of the stochastic process $\tilde{I}_t^{a^{n*}}$ in (6.13), $\tilde{I}_t^{a^{n*}}$ refreshes when $t = \tau_n$. As a result, the distribution of $\tilde{I}_{\tau_n}^{a^{n*}}$ is independent of the event $\mathcal{B}_{n-1}$ and is identical to the distribution of $I_\infty^{a^{n*}}$, which implies that

$$P(\tilde{I}_{\tau_n}^{a^{n*}} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}_{\tau_n}^{a^{n*}} \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2} \tag{E.6}$$

where the second inequality follows from Lemma E.0.4. Moreover, conditioning on $\mathcal{B}_{n-1}$, since

I_{τ_n-1} couples with $\tilde{I}^{a^{(n-1)*}}_{\tau_n-1}$, we have that

$$P(I_{\tau_n-1} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \leq \kappa_1 \cdot \log T)/P(\mathcal{B}_{n-1}) \quad \text{(E.7)}$$

Note that the distribution of $\tilde{I}^{a^{(n-1)*}}_{\tau_n-1}$ is identical to the distribution of $I^{a^{(n-1)*}}_{\infty}$, which implies that

$$P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \leq \kappa_1 \cdot \log T) = P(I^{a^{(n-1)*}}_{\infty} \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2}$$

Therefore, by noting that $P(\mathcal{B}_{n-1}) \leq 1$, from (E.7), we have that

$$P(I_{\tau_n-1} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \geq 1 - \frac{1}{T^2} \quad \text{(E.8)}$$

From (E.6), (E.8) and the union bound, we have that

$$P(I_{\tau_n-1} \leq \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \geq 1 - \frac{2}{T^2} \quad \text{(E.9)}$$

As a result, conditioning on $\mathcal{B}_{n-1}$, we know that

$$I_{\tau_n+L} \leq L \cdot \bar{D} + \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n+L} \leq L \cdot \bar{D} + \kappa_1 \cdot \log T \quad \text{(E.10)}$$

happens with a probability at least $1 - \frac{2}{T^2}$. It is clear to see that the event $\{I_{\tau_n+\max\{\log T, 2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\max\{\log T, 2L\}}\}$ happens as long as

$$I_t = \tilde{I}^{a^{n*}}_t = 0 \quad \text{(E.11)}$$

for some $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$.

Suppose that $I_{\tau_n} \leq \tilde{I}^{a^{n*}}_{\tau_n}$ (resp. $I_{\tau_n} \geq \tilde{I}^{a^{n*}}_{\tau_n}$), from the evolution of the stochastic process in (6.10) and (6.13), we have that $I_t \leq \tilde{I}^{a^{n*}}_t$ (resp. $I_t \geq \tilde{I}^{a^{n*}}_t$) for any $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$. Given that I_t and $\tilde{I}^{a^{n*}}_t$ must be non-negative (from definition), we conclude that if $I_{\tau_n} \leq \tilde{I}^{a^{n*}}_{\tau_n}$ (resp.

$I_{\tau_n} \geq \tilde{I}^{a^{n*}}_{\tau_n}$), then (E.11) happens as long as $\tilde{I}^{a^{n*}}_t = 0$ (resp. $I_t = 0$). Thus, a sufficient condition for (E.11) to happen is that

$$\sum_{t=\tau_n+L}^{\tau_n+\max\{\log T,2L\}} D_t - s(a^{n*}, Z_t) \geq L \cdot \bar{D} + \kappa_1 \cdot \log T$$

Since $D_t - s(a^{n*}, Z_t)$ are i.i.d. random variable for $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$, we denote by $\delta_n = \mathbb{E}_{D\sim F}[D] - \mathbb{E}_{Z\sim G}[s(a^{n*}, Z)] \geq \delta$. Following Hoeffding's inequality (Lemma E.0.1), we have that

$$\begin{aligned} P\left(\sum_{t=\tau_n+L}^{\tau_n+\kappa_2\max\{\log T,2L\}} D_t - s(a^{n*}, Z_t) \geq L \cdot \bar{D} + \kappa_1 \cdot \log T\right) &\geq 1 - \exp\left(-\frac{2(\kappa_2 \max\{\log T, 2L\} - L(\bar{D}+1) - \kappa_1 \log T)^2}{\kappa_2 \max\{\log T, 2L\} - L}\right) \\ &\geq 1 - \frac{1}{T^2} \end{aligned}$$

where $\kappa_2 \geq \max\{4, 2(\bar{D}+1+\kappa_1)\} \geq \max\{\frac{4\log T}{\max\{\log T,2L\}}, 2(\bar{D}+1+\kappa_1)\}$. Therefore, we have that

$$P\left(I_{\tau_n+\kappa_2\max\{\log T,2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\kappa_2\max\{\log T,2L\}} \mid \mathcal{B}_{n-1} \text{ and (E.10) happens}\right) \geq 1 - \frac{1}{T^2}.$$

Combining (E.9) and the induction hypothesis that $P(\mathcal{B}_{n-1}) \geq 1 - \frac{3(n-1)}{T^2}$, we have that

$$\begin{aligned} P(\mathcal{B}_n) =& P(\mathcal{B}_{n-1}) \cdot P(I_{\tau_n-1} \leq \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n} \leq \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \\ & \cdot P\left(I_{\tau_n+\kappa_2\max\{\log T,2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\kappa_2\max\{\log T,2L\}} \mid \mathcal{B}_{n-1} \text{ and (E.10) happens}\right) \\ \geq& (1 - \frac{3(n-1)}{T^2}) \cdot (1 - \frac{2}{T^2}) \cdot (1 - \frac{1}{T^2}) \geq (1 - \frac{3(n-1)}{T^2}) \cdot (1 - \frac{3}{T^2}) \\ \geq& 1 - \frac{3n}{T^2} \end{aligned}$$

which completes our proof of the induction of (E.4) for each epoch n. Therefore, our proof of the lemma is completed. □

Proof of Lemma 6.3.4. Clearly, from (6.14), it is enough to compare the value of I_t and $\tilde{I}_t^{a^{n*}}$ for each epoch n and each period t in the epoch n. Note that we identify an event $\mathcal{B}$ in Lemma 6.3.3 that I_t and $\tilde{I}_t^{a^{n*}}$ couple with each other. We consider two situations where $\mathcal{B}$ happens or $\mathcal{B}$ not happens.

Case 1: We now assume that $\mathcal{B}$ happens. Then, we know that for each epoch $n \in [N]$ and each $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1$, the value of I_t and $\tilde{I}_t^{a^{n*}}$ are identical. Therefore, only when $t = \tau_n, \ldots, \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}$, the value of I_t and $\tilde{I}_t^{a^{n*}}$ can be different. Moreover, note that the evolution of I_t in (6.10) is the same as the evolution of $\tilde{I}_t^{a^{n*}}$ in (6.13), except that the intial value I_{τ_n} is different from $\tilde{I}_{\tau_n}^{a^{n*}}$. We know that the gap between I_t and $\tilde{I}_t^{a^{n*}}$ can only become smaller. Therefore, we get that

$$|I_t - \tilde{I}_t^{a^{n*}}| \leq |I_{\tau_n} - \tilde{I}_{\tau_n}^{a^{n*}}| \leq \kappa_1 \cdot \log T \tag{E.12}$$

where the last inequality follows from the condition in the event $\mathcal{B}$. We have that

$$\begin{aligned} \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B} \right] \right| &\leq \sum_n \sum_{t=\tau_n}^{\tau_n + \kappa_2 \cdot \max\{\log T, 2L\}} \mathbb{E}[|I_t - \tilde{I}_t^{a^{n*}}| \mid \mathcal{B}] \\ &\leq \sum_n \kappa_1 \cdot \log T \cdot \kappa_2 \cdot \max\{\log T, 2L\} \\ &= N \cdot \kappa_1 \kappa_2 \log T \cdot \max\{\log T, 2L\} \end{aligned} \tag{E.13}$$

Case 2: We now assume that $\mathcal{B}$ does not happen. Clearly, a direct upper bound on both I_t and $\tilde{I}_t^{a^{n*}}$ is that

$$I_t \leq \bar{D} \cdot t \text{ and } \mathbb{E}[\tilde{I}_t^{a^{n*}} \mid \mathcal{B}^c] \leq \bar{D} \cdot t$$

Therefore, we have that

$$\left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c \right] \right| \leq \bar{D} \cdot \sum_{t=1}^{T} t \leq \bar{D} \cdot T^2 \tag{E.14}$$

However, from Lemma 6.3.3, we know that $P(\mathcal{B}^c) \leq \frac{3N}{T^2}$. As a result, combining (E.13) and (E.14), we get that

$$\begin{aligned}
\left|\mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}})\right]\right| &\leq \left|\mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}\right]\right| \cdot P(\mathcal{B}) + \left|\mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c\right]\right| \cdot P(\mathcal{B}^c) \\
&\leq \left|\mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}\right]\right| + \left|\mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c\right]\right| \cdot \frac{3N}{T^2} \\
&\leq N \cdot \kappa_1\kappa_2 \log T \cdot \max\{\log T, 2L\} + 3N\bar{D}
\end{aligned}$$

which completes our proof. □

Proof of Lemma 6.3.5. The proof generalizes the proof of Lemma 6.3.3. For each epoch n, we denote by

$$C_n = \{I^a_{\tau_{n'}} \leq \kappa_1 \cdot \log T,\ \tilde{I}^a_{\tau_{n'}} \leq \kappa_1 \cdot \log T \text{ and } I^a_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{n'*}}_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}}\},\ \forall a \in \mathcal{A}_{n'},\ \forall n' \leq n\}.$$

In order to prove the lemma, it is sufficient to prove that

$$P(C_n) \geq 1 - \frac{3(K+1)n}{T^2}. \tag{E.15}$$

We prove (E.15) by using induction on the epoch n.

Clearly, when $n = 1$, we have that $P(I^a_{\tau_1} = 0) = 1$ for all $a \in \mathcal{A}_1$. From Lemma E.0.4, there exists a constant $\kappa_1 > 0$ such that for each $a \in \mathcal{A}_1$, it holds that

$$P(\tilde{I}^a_{\tau_1} \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2}$$

by noting that the distribution of $\tilde{I}^a_{\tau_1}$ is identical to the distribution of I^a_∞ for each $a \in \mathcal{A}_1$.

Now conditioning on the event $\{\tilde{I}^a_{\tau_1} \leq \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1\}$, we proceed to bound the probability that event $\{I^a_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^a_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}},\ \forall a \in \mathcal{A}_1\}$ happens. Note that the evolution of

the stochastic process I_t^a in (6.9) is identical to the evolution of the stochastic process $\tilde{I}_t^a$ in (6.15) for $t = \tau_1, \ldots, \tau_2 - 1$. Therefore, for each $a \in \mathcal{A}_1$, it is clear to see that the event $\{I^a_{\tau_1+\kappa_2\cdot\max\{\log T, 2L\}} = \tilde{I}^a_{\tau_1+\kappa_2\cdot\max\{\log T, 2L\}}\}$ happens as long as

$$I_t^a = \tilde{I}_t^a = 0, \quad \text{for some } t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}. \tag{E.16}$$

We note that $I^a_{\tau_1} \leq \tilde{I}^a_{\tau_1}$ implies that $I_t^a \leq \tilde{I}_t^a$ for all $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. From the non-negativity of I_t^a and $\tilde{I}_t^a$, we have that (E.16) holds as long as $\tilde{I}_t^a = 0$ for some $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. As a result, a sufficient condition for (E.16) to hold for a $a \in \mathcal{A}_1$ is that

$$\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \geq \kappa_1 \cdot \log T$$

Note that $D_t - s(a, Z_t)$ are i.i.d. random varibles for $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. We also denote by $\delta = \mathbb{E}[D] - \mathbb{E}[s(\bar{q}, Z)]$. Following Hoeffding's inequality (Lemma E.0.1), we have that

$$P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \geq \kappa_1 \cdot \log T\right) \geq 1 - \exp\left(-\frac{2(\delta\kappa_2 \max\{\log T, 2L\} - \kappa_1 \log T)^2}{\kappa_2 \cdot \max\{\log T, 2L\} \cdot \bar{D}}\right) \geq 1 - \frac{1}{T^2}$$

where $\kappa_2 \geq \max\{\frac{2\kappa_1}{\delta}, \frac{4\bar{D}}{\delta^2}\} \geq \max\{\frac{2\kappa_1 \log T}{\delta \cdot \max\{\log T, 2L\}}, \frac{4\bar{D}\log T}{\delta^2 \cdot \max\{\log T, 2L\}}\}$.

The above derivation implies that

$$\begin{aligned}
P(\mathcal{B}_1) &= P\left(\mathcal{B}_1 \mid \tilde{I}^a_{\tau_1} \leq \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_1\right) \cdot P(\tilde{I}^a_{\tau_1} \leq \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_1) \\
&= P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \geq \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_1\right) \cdot P(\tilde{I}^{a^{1*}}_{\tau_1} \leq \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_1) \\
&\geq (1 - \frac{K+1}{T^2}) \cdot (1 - \frac{K+1}{T^2}) \geq 1 - \frac{2(K+1)}{T^2} \geq 1 - \frac{3(K+1)}{T^2}
\end{aligned}$$

where the first inequality follows from the union bound by noting that $|\mathcal{A}_1| \leq K + 1$. Therefore, we prove (E.15) for $n = 1$.

Now assume that (E.15) holds for epoch $n-1$. We consider epoch n. Clearly, from the definition of the stochastic process $\tilde{I}_t^a$ in (6.15), $\tilde{I}_t^a$ refreshes when $t = \tau_n$. As a result, the distribution of $\tilde{I}_{\tau_n}^a$ is independent of the event C_{n-1} and is identical to the distribution of I_∞^a, which implies that

$$P(\tilde{I}_{\tau_n}^a \leq \kappa_1 \cdot \log T \mid C_{n-1}) = P(\tilde{I}_{\tau_n}^a \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2}, \quad \forall a \in \mathcal{A}_n \tag{E.17}$$

where the second inequality follows from Lemma E.0.4. Moreover, conditioning on $\mathcal{B}_{n-1}$, since $I_{\tau_n-1}^a$ couples with $\tilde{I}_{\tau_n-1}^a$ for each $a \in \mathcal{A}_n \subset \mathcal{A}_{n-1}$, we have that

$$P(I_{\tau_n-1}^a \leq \kappa_1 \cdot \log T \mid C_{n-1}) = P(\tilde{I}_{\tau_n-1}^a \leq \kappa_1 \cdot \log T \mid C_{n-1}) = P(\tilde{I}_{\tau_n-1}^a \leq \kappa_1 \cdot \log T)/P(C_{n-1}), \ \forall a \in \mathcal{A}_n \tag{E.18}$$

Note that the distribution of $\tilde{I}_{\tau_n-1}^a$ is identical to the distribution of I_∞^a, which implies that

$$P(\tilde{I}_{\tau_n-1}^a \leq \kappa_1 \cdot \log T) = P(I_\infty^a \leq \kappa_1 \cdot \log T) \geq 1 - \frac{1}{T^2}, \ \forall a \in \mathcal{A}_n$$

Therefore, by noting that $P(C_{n-1}) \leq 1$, from (E.18), we have that

$$P(I_{\tau_n-1}^a \leq \kappa_1 \cdot \log T \mid C_{n-1}) \geq 1 - \frac{1}{T^2}, \ \forall a \in \mathcal{A}_n. \tag{E.19}$$

From (E.17), (E.19) and the union bound, we have that

$$P(I_{\tau_n-1} \leq \kappa_1 \cdot \log T \text{ and } \tilde{I}_{\tau_n}^{a^{n*}} \leq \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_n \mid C_{n-1}) \geq 1 - \frac{2(K+1)}{T^2} \tag{E.20}$$

where we note that $|\mathcal{A}_n| \leq K + 1$. As a result, conditioning on C_{n-1}, we know that

$$I_{\tau_n+L}^a \leq L \cdot \bar{D} + \kappa_1 \cdot \log T \text{ and } \tilde{I}_{\tau_n+L}^a \leq L \cdot \bar{D} + \kappa_1 \cdot \log T, \ \forall a \in \mathcal{A}_n \tag{E.21}$$

happens with a probability at least $1 - \frac{2(K+1)}{T^2}$. It is clear to see that for each $a \in \mathcal{A}_n$, the event

$\{I^a_{\tau_n+\max\{\log T,2L\}} = \tilde{I}^a_{\tau_n+\max\{\log T,2L\}}\}$ happens as long as

$$I^a_t = \tilde{I}^a_t = 0 \tag{E.22}$$

for some $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$.

For each $a \in \mathcal{A}_n$, suppose that $I^a_{\tau_n} \leq \tilde{I}^a_{\tau_n}$ (resp. $I^a_{\tau_n} \geq \tilde{I}^a_{\tau_n}$), from the evolution of the stochastic process in (6.9) and (6.15), we have that $I^a_t \leq \tilde{I}^a_t$ (resp. $I^a_t \geq \tilde{I}^a_t$) for any $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$. Given that I^a_t and $\tilde{I}^a_t$ must be non-negative (from definition), we conclude that if $I^a_{\tau_n} \leq \tilde{I}^a_{\tau_n}$ (resp. $I^a_{\tau_n} \geq \tilde{I}^a_{\tau_n}$), then (E.11) happens as long as $\tilde{I}^a_t = 0$ (resp. $I^a_t = 0$). Thus, a sufficient condition for (E.22) to happen is that

$$\sum_{t=\tau_n+L}^{\tau_n+\max\{\log T,2L\}} D_t - s(a, Z_t) \geq L \cdot \bar{D} + \kappa_1 \cdot \log T$$

Since $D_t - s(a, Z_t)$ are i.i.d. random variable for $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$, we denote by $\delta_{n,a} = \mathbb{E}_{D\sim F}[D] - \mathbb{E}_{Z\sim G}[s(a, Z)] \geq \delta$. Following Hoeffding's inequality (Lemma E.0.1), for each $a \in \mathcal{A}_n$, we have that

$$\begin{aligned} P\left(\sum_{t=\tau_n+L}^{\tau_n+\kappa_2\max\{\log T,2L\}} D_t - s(a, Z_t) \geq L \cdot \bar{D} + \kappa_1 \cdot \log T\right) &\geq 1 - \exp\left(-\frac{2(\kappa_2 \max\{\log T, 2L\} - L(\bar{D}+1) - \kappa_1 \log T)^2}{\kappa_2 \max\{\log T, 2L\} - L}\right) \\ &\geq 1 - \frac{1}{T^2} \end{aligned}$$

where $\kappa_2 \geq \max\{4, 2(\bar{D} + 1 + \kappa_1)\} \geq \max\{\frac{4\log T}{\max\{\log T,2L\}}, 2(\bar{D} + 1 + \kappa_1)\}$. Therefore, from the union bound, we have that

$$P\left(I^a_{\tau_n+\kappa_2\max\{\log T,2L\}} = \tilde{I}^a_{\tau_n+\kappa_2\max\{\log T,2L\}},\ \forall a \in \mathcal{A}_n \mid C_{n-1} \text{ and (E.21) happens}\right) \geq 1 - \frac{K+1}{T^2}.$$

Combining (E.20) and the induction hypothesis that $P(C_{n-1}) \geq 1 - \frac{3(K+1)(n-1)}{T^2}$, we have that

$$\begin{aligned}
P(C_n) =& P(C_{n-1}) \cdot P(I^a_{\tau_n - 1} \leq \kappa_1 \cdot \log T \text{ and } \tilde{I}^a_{\tau_n} \leq \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_n \mid C_{n-1}) \\
& \cdot P\left(I^a_{\tau_n + \kappa_2 \max\{\log T, 2L\}} = \tilde{I}^a_{\tau_n + \kappa_2 \max\{\log T, 2L\}},\ \forall a \in \mathcal{A}_n \mid C_{n-1} \text{ and (E.21) happens}\right) \\
\geq & (1 - \frac{3(K+1)(n-1)}{T^2}) \cdot (1 - \frac{2(K+1)}{T^2}) \cdot (1 - \frac{K+1}{T^2}) \\
\geq & (1 - \frac{3(K+1)(n-1)}{T^2}) \cdot (1 - \frac{3(K+1)}{T^2}) \\
\geq & 1 - \frac{3(K+1)n}{T^2}
\end{aligned}$$

which completes our proof of the induction of (E.15) for each epoch n. Therefore, our proof of the lemma is completed. □

Proof of Lemma 6.3.6. We first show that for each epoch $n \in [N]$ and each action $a \in \mathcal{A}_n$, we can use the average value of $\tilde{I}^a_t$ for $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}$ to $\tau_{n+1} - 1$ to approximate the value of $\mathbb{E}[I^{\pi_a}_\infty]$, where the length of the confidence interval can be given by γ_n.

Clearly, $\{\tilde{I}^a_t\}^{\tau_{n+1}-1}_{t=\tau_n + \kappa_2 \cdot \max\{\log T, 2L\}}$ forms a Markov chain. We denote by $\boldsymbol{I}$ a vector such that

$$\boldsymbol{I} = (\tilde{I}^a_t, \forall t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1)$$

We apply Lemma E.0.2 to derive a concentration bound for $\boldsymbol{I}$. To be specific, for each epoch $n \leq N-1$, we regard $\tilde{I}^a_{\tau_n + \kappa_2 \cdot \max\{\log T, 2L\} + i}$ as X_i for $i = 1, \ldots, \tau_{n+1} - 1 - \tau_n - \kappa_2 \cdot \max\{\log T, 2L\}$. Clearly, $\boldsymbol{I}$ is a Markov chain with stationary distributions and satisfies the conditions in Lemma E.0.2.

We now denote $m = \tau_{n+1} - \tau_n - 1 - \kappa_2 \cdot \max\{\log T, 2L\}$. Then, from Lemma E.0.2, there exists a constant λ such that for any $\epsilon > 0$, it holds

$$P\left(\left|\sum_{i=1}^{m} \tilde{I}^a_{\tau_n + \kappa_2 \cdot \max\{\log T, 2L\} + i} - \mathbb{E}\left[\sum_{i=1}^{m} \tilde{I}^a_{\tau_n + \kappa_2 \cdot \max\{\log T, 2L\} + i}\right]\right| \geq \sqrt{m} \cdot \epsilon\right) \leq 2 \exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right).$$

We now set $\epsilon = \frac{\gamma_n \cdot \sqrt{m}}{2}$. Then, we have

$$\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \leq \exp(-\frac{(1-\lambda)m\gamma_n^2}{16}) \tag{E.23}$$

We proceed to give a lower bound on $m\gamma_n^2$, which will imply an upper bound for (E.23). Note that

$$m = \tau_{n+1} - \tau_n - 1 - \kappa_2 \cdot \max\{\log T, 2L\} = \kappa_2 \cdot (\max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} - \max\{\log T, 2L\})$$

If $\frac{1}{\gamma_2} \cdot \log T \geq 3L$, then we have

$$\max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} - \max\{\log T, 2L\} = \frac{1}{\gamma_n^2} \cdot \log T - \max\{\log T, 2L\} \geq \frac{1}{3\gamma_n^2} \cdot \log T$$

If $\frac{1}{\gamma_2} \cdot \log T < 3L$, then we have

$$\max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} - \max\{\log T, 2L\} = L \geq \frac{1}{3\gamma_n^2} \cdot \log T$$

Therefore, it holds that

$$m = \kappa_2 \cdot (\max\{\frac{1}{\gamma_n^2} \cdot \log T, 3L\} - \max\{\log T, 2L\}) \geq \frac{\kappa_2}{3\gamma_n^2} \cdot \log T \tag{E.24}$$

Plugging (E.24) into (E.23), we have

$$\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \leq \exp\left(-\frac{(1-\lambda)\kappa_2 \log T}{12}\right) \leq \frac{1}{T^2}$$

where $\kappa_2 \geq 24/(1-\lambda)$. We have

$$\begin{aligned} & P\left(\left|\sum_{i=1}^{m} \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i} - \mathbb{E}\left[\sum_{i=1}^{m} \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i}\right]\right| \geq m\cdot\frac{\gamma_n}{2}\right) \\ =& P\left(\left|\sum_{i=1}^{m} \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i} - m\cdot\mathbb{E}\left[I^{\pi_a}_{\infty}\right]\right| \geq m\cdot\frac{\gamma_n}{2}\right) \\ \leq& \frac{2}{T^2} \end{aligned} \tag{E.25}$$

where the second inequality follows from the fact that the distribution of $\tilde{I}^a_t$ is identical to the distribution of $I^{\pi_a}_{\infty}$. Moreover, from Hoeffding's inequality (Lemma E.0.1), it holds that

$$\begin{aligned} P\left(\left|\sum_{t=\tau_n+\kappa_2\cdot\max\{\log T,2L\}}^{\tau_{n+1}-1} s(a,Z_t)) - m\cdot\mathbb{E}[s(a,Z)]\right| \geq m\cdot\frac{\gamma_n}{2}\right) &\leq 2\exp(-\frac{m\gamma_n^2}{2\bar{D}^2}) \leq 2\exp(-\frac{\kappa_2\log T}{6\bar{D}^2}) \\ &\leq \frac{2}{T^2} \end{aligned} \tag{E.26}$$

where the second inequality follows from (E.24) and the third inequality follows from $\kappa_2 \geq 12\bar{D}^2$. Therefore, conditional on the event C happens, we have that

$$P\left(\left|\tilde{C}^a_n - \hat{C}^{\pi_a}_{\infty}\right| \leq (h+b)\cdot\frac{\gamma_n}{2} \mid C\right) \geq 1 - \frac{4}{T^2}$$

which implies that (from union bound over all $a \in \mathcal{A}$ and all $n \leq N-1$)

$$P(\mathcal{E} \mid C) = P\left(\{|\tilde{C}^a_n - \hat{C}^{\pi_a}_{\infty}| \leq (h+b)\cdot\frac{\gamma_n}{2},\ \forall a \in \mathcal{A}_n, \forall 1 \leq n \leq N-1\}\right) \geq 1 - \frac{4(K+1)N}{T^2}$$

From Lemma 6.3.5, we know that $P(C) \geq 1 - \frac{3(K+1)N}{T^2}$. Therefore, we have that

$$P(\mathcal{E}) = P(\mathcal{E} \mid C)\cdot P(C) \geq 1 - \frac{7(K+1)N}{T^2}$$

which completes our proof. □