Optimal and Near-optimal Policies of Sequential Decision Problems under Uncertainty

by

Xinchang Xie

Business Administration
Duke University

Date: _____

Approved:

_____
Alessandro Arlotto, Advisor

_____
Alexandre Belloni, Co-Advisor

_____
Yehua Wei

_____
Shankar Bhamidi

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Business Administration
in the Graduate School of
Duke University

2020

ABSTRACT

Optimal and Near-optimal Policies of Sequential Decision Problems under Uncertainty

by

Xinchang Xie

Business Administration
Duke University

Date: _____

Approved:

_____
Alessandro Arlotto, Advisor

_____
Alexandre Belloni, Co-Advisor

_____
Yehua Wei

_____
Shankar Bhamidi

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2020

Abstract

In this dissertation, we study stochastic sequential decision problems with a discrete finite time horizon. We conduct substantial analyses on both optimal and near-optimal policies of such problems, and focus on the *total reward* collected by different sequential policies from different perspectives. The total reward collected by any sequential policy is *a priori* random, and for many of the policies we study, we are interested in: (i) the expected total reward they collect, (ii) the variance of their total reward, (iii) the limiting distribution of their total reward, and (iv) confidence intervals for their total reward. With these goals in mind, we consider different sequential decision problems that include the dynamic and stochastic knapsack problem with equal rewards, the problem of sequential monotone subsequence selection from a random sample, and a dynamic inventory control problem. For the dynamic and stochastic knapsack problem with equal rewards, we propose an adaptive heuristic policy and prove that its regret—the expected gap between the total reward collected by the proposed heuristic policy and that collected by the best offline solution—is at most logarithmically in the problem size. Furthermore, we characterize the variance asymptotics as well as the limiting distribution of the total reward collected by such heuristic. We show that the total reward collected by our heuristic policy shares the same variance asymptotics and the same limiting distribution with that collected by the optimal sequential policy. This is in contrast with the performance of other asymptotically optimal heuristics whose total rewards have larger regrets, larger variances, and different limiting distributions. We also discuss the equivalence between the problem of sequential monotone subsequence selection from a random sample and a special instance of the dynamic and stochastic knapsack problem with equal rewards. Such equivalence enables us to apply the above mentioned analyses and results to the sequential monotone subsequence selection problem, and establish a number of results of independent interest. Lastly, we study multiple finite horizon sequential decision problems that are faced in parallel by self-interested decision makers. The uncertainties they face could be correlated in an arbitrary unknown fashion. Our goal is to construct simultaneous confidence intervals for the total rewards collected by each decision maker, through the observations of the rewards they collect along the time horizon. We propose a data-driven bootstrap procedure that is asymptotically valid. The key feature of such procedure is that the number of decision makers is allowed to be much larger than the length of the time horizon. This

corresponds to the high-dimensional regime in the statistics literature in which the dimension of the estimator is much larger than sample size. As a byproduct, such bootstrap procedure can be used to simultaneously test whether each of the decision makers is implementing certain specified policy. We use a dynamic inventory control problem and the sequential monotone subsequence selection problem to illustrate the effectiveness of our framework.

Acknowledgments

The Ph.D. process has been a challenging yet enjoyable journey for me, and I could not have completed it without my co-advisors Prof. Alessandro Arlotto and Prof. Alex Belloni.

It is my great fortune to have Alessandro as my co-advisor, and I could not have made it to accomplish this milestone without his guidance and support. Through our weekly meetings since my second year of the program, I have learned from him not only how to conduct high standard research, but also how to (and how important it is to) articulate research results in papers and presentations. He has also been guiding me with patience and encouragement, through which I grew—step by step—as a researcher and as a person.

I am also deeply thankful to Alex for serving my other co-advisor. With his broad research interests, knowledge, and expertise, Alex has taught me how to connect different fields to find out interesting research questions. He has been truly supportive and would always make time out of his busy schedule to address my questions. He has also been leading by example and teaching me the importance of hardworking to succeed.

My sincere thanks also go to the other two professors on my committee: Prof. Yehua Wei, who was my mentor during my first year and has been supporting me since then, and Prof. Shankar Bhamidi, who kindly served on my committee as an external member and has provided valuable comments and perspectives for my candidacy exam, proposal exam, and my final defense. In addition, I want to thank several other faculty members at Fuqua: Peng Sun, David Brown, Bob Nau, Bob Winkler, and Bora Keskin. Through their courses, I deepened my knowledge and learned tools that have been helpful for my research.

Besides all the help I got from the faculty, family and friends have been supporting me since the very beginning. My deepest appreciation is for my parents, who have been supporting me unconditionally all the time. I would also like to thank my dear friends within the Decision Sciences area and the Operations Management area at Fuqua: Levi DeValve, Huseyin Gurkan, Mingliu Chen, Chen Chen, Yunke Mai, and Soudipta Chakraborty–with whom I took classes with, and discussed research problems through the years. More importantly, I am grateful for the many Fuqua Fridays that we spent together. I am also thankful for my other friends across different departments at

Duke: Yanyou Chen, Joy Tong, Anyi Ma, Wenxi Liao, Ye Jin, Xuyan Ru, David Hall, and Honggi Lee. It is you guys who made this journey more enjoyable and unforgettable.

Finally, I am gratefully acknowledging partial financial support from the National Science Foundation (Grant No. 1553274).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this dissertation, we study finite-horizon sequential decision problems under uncertainty, and analyze their optimal and near-optimal policies. Sequential decision problems are widely used in operations management and operations research; their versatility can be easily seen through a wide range of applications. The finite horizon formulation of such problems can be typically described as follows. In each time period, a decision maker (referred to as *she*) sees the realization of an exogenous stochastic process and has to decide what action to take based on the available information and on her past actions. The realized uncertainty and the chosen action then determine a reward she collects during that time period. The objective is typically to maximize the expected total collected reward. The optimal policies of such problems are usually difficult to analyze as that they are both time- and state-dependent. While one can in principle analyze such dependencies by solving the problems to optimality through the associated Bellman equations, such solutions are often lack of closed-form expressions, and their computations suffer the infamous *curse of dimensionality* (Bellman, 1957; Powell, 2011).

As a substitute, decision makers often seek for near-optimal heuristics with provable performance guarantees. Such heuristics are usually evaluated in terms of *optimality gap* or *regret*. Optimality gap is the expected performance gap between a heuristic and the optimal sequential policy. Instead, regret is the expected performance gap between a heuristic and the optimal offline policy. Closely related to these two concepts is the notion of *asymptotic optimality*. A heuristic sequential policy is said to be asymptotically optimal if the ratio between its optimality gap and the expected performance of the optimal sequential policy goes to zero as the problem size goes to infinity. In this dissertation, we will evaluate different policies for certain problems in terms of these metrics.

In addition to these metrics that are based on the expected performance, we also study sequential decision problems from a distributional standpoint. We analyze the variance of the total reward, which is critical to differentiate policies that have asymptotically equivalent means. We also analyze of the limiting distribution of the total reward, which is helpful for constructing confidence intervals for the total reward. All such analyses have both theoretical and practical benefits.

From a theoretical standpoint, such analyses would help us further differentiate different policies within the class of asymptotically optimal policies. From a practical standpoint, the construction of confidence intervals could be applied to compute the value at risk of different policies and to test whether decision makers are implementing certain policies through the observations of the rewards they collect over time.

The sequential decision problems we will cover in this dissertation include the dynamic and stochastic knapsack problem with equal rewards, the problem of sequential monotone subsequence selection from a random sample, and a dynamic inventory control problem. In Chapter 2, we study the dynamic and stochastic knapsack problem with equal rewards. Such a formulation of knapsack problem was studied by Coffman et al. (1987) and Kleywegt and Papastavrou (1998, Section 5). For this problem, we propose and study a simple adaptive heuristic with closed-form expression, and prove that its regret—the expected performance difference between such heuristic and the offline optimal solution—grows at most logarithmically in the problem size. In Chapter 3, we discuss how the logarithmic regret bound of Chapter 2 carries over to the equivalent problem of sequential selection of an increasing subsequence from a random sample.

Then in Chapter 4, we go back to the dynamic and stochastic knapsack problem with equal rewards and offer a distributional analysis. Specifically, we focus on the higher-order performance and limiting behavior of the heuristic proposed in Chapter 2, and prove that it is not only asymptotically optimal, but also that it collects a total reward that has the same variance asymptotics and the same limiting distribution as that collected by the optimal sequential policy. We note that the total reward collected by other asymptotically optimal heuristic studied in the literature may have different (larger) variance and a different limiting distribution than that collected by the optimal sequential policy.

When one has the available limiting distribution of the total reward collected by a policy for a sequential decision problem, then it is easy to construct confidence intervals for the total reward, or to conduct hypothesis testing of whether the decision maker is implementing such a policy. In Chapter 5, we consider such two statistical tasks when there are many sequential decision problems faced by self-interested decision makers and the uncertainties they face are correlated in an arbitrary unknown fashion. To this end, we propose a bootstrap based procedure for constructing simultaneous confidence intervals for the total rewards collected by each decision maker. As a key

feature, such procedure allows for the number of decision makers being comparable to or even much larger than the length of the time horizon. This regime is often referred to as *high dimensional*; the the number of decision makers can be viewed as the data's dimension and the length of the time horizon can be interpreted as sample size. Lastly, we use a dynamic inventory control problem and the sequential monotone subsequence selection problem to illustrate the effectiveness of the framework.

Chapter 2

Logarithmic Regret in the Dynamic and Stochastic Knapsack Problem with Equal Rewards

The knapsack problem is one of the classic problems in operations research. It arises in resource allocation, and it counts numerous applications in auctions, logistics, portfolio optimization, scheduling, and transportation among others (cf. Martello and Toth, 1990; Kellerer et al., 2004). In its dynamic and stochastic formulation (see, e.g. Papastavrou et al., 1996; Kleywegt and Papastavrou, 1998, 2001) a decision maker (referred to as *she*) is given a knapsack with finite capacity $0 \leqslant c < \infty$ and is sequentially presented with items arriving over a time horizon with n discrete time periods, indexed by $i \in [n] \equiv \{1, 2, \ldots, n\}$. In each period $i \in [n]$, an item arrives with probability p, its weight-reward pair $(\mathfrak{W}_i, \mathfrak{R}_i)$ is revealed, and the decision maker needs to decide whether to include the arriving item in the knapsack or to reject it forever. Here, the weight $\mathfrak{W}_i$ represents the amount of knapsack capacity that the item arriving in period i consumes if the decision maker chooses to include it in the knapsack, and the reward $\mathfrak{R}_i$ represents what the decision maker collects upon inclusion. The pairs $(\mathfrak{W}_i, \mathfrak{R}_i)$, $i \in [n]$, are independent and with common, known, bivariate distribution supported on the nonnegative orthant.

By imposing different assumptions on the weight-reward distribution, one recovers knapsack instances of independent interest. For instance, in the problem of real-time uniprocessor scheduling under conditions of overload (see, e.g., Baruah et al., 1994) a decision maker wants to maximize the number of jobs that are processed on a single machine by a fixed deadline. In this context, the deadline is the knapsack capacity and jobs correspond to items. Their rewards are all equal to one, and their durations correspond to the item weights. This scheduling application motivates the model in this chapter. We assume that the rewards are deterministic and all equal[1] to $r > 0$, and the weights are independent random variables with common continuous distribution F. We model item arrivals by considering a Bernoulli process $\mathfrak{B}_1, \mathfrak{B}_2, \ldots, \mathfrak{B}_n$ that is independent of everything else, and that is given by a sequence of independent Bernoulli random variables with success probability

This chapter is written under the supervision of Prof. Alessandro Arlotto. The results presented here are also in the joint paper Arlotto and Xie (2020a), forthcoming at *Stochastic Systems*.

[1]This also covers random rewards with common distribution that are revealed only after the inclusion decision.

p. We then equivalently redefine the weight distribution so that a no arrival corresponds to the arrival of an item with arbitrarily large weight. That is, we assume that an item arrives in each period $i \in [n]$ and that its weight is given by the random variable W_i defined by

$$W_i = \begin{cases} +\infty & \text{if } \mathfrak{B}_i = 0 \\ \mathfrak{W}_i & \text{if } \mathfrak{B}_i = 1. \end{cases}$$

We say that a policy π is *feasible* if the sum of the weights of the items selected by π does not exceed the knapsack capacity c, and we say that the policy is *online* (or *sequential*) if the decision to select item i with weight W_i depends only on the information available up to and including time i. We then let $\Pi(n, c, p)$ be the set of feasible online policies, and we compare the performance of the best online policy to that of a prophet who has full (or *offline*) knowledge of the weights $W_1, W_2, \ldots, W_n$ before making any selection. Under some mild technical conditions on the weight distribution F, we prove that the regret—the expected gap between the performance of the best online policy and its offline counterpart—is bounded by the logarithm of n. Our proof is constructive. We propose a reoptimized heuristic that exhibits logarithmic regret. The heuristic is based on resolving some related optimization problem at any given time $i \in [n]$ by using the current—rather than the initial—level of remaining capacity as constraint. The solution of this optimization problem provides us with a state- and time-dependent threshold that mimics that of the optimal online policy.

If all of the weights $W_1, W_2, \ldots, W_n$ are revealed to the decision maker before she makes any selection, then her choice is obvious. To maximize the total reward she collects, she just sorts the items according to their weights and selects them starting from the smallest weight and continuing until the knapsack capacity is exhausted. Formally, if $W_{(1,n)} \leqslant W_{(2,n)} \leqslant \cdots \leqslant W_{(n,n)}$ are the order statistics of $W_1, W_2, \ldots, W_n$, then the maximal reward $R_n^*(c, p, r)$ that the decision maker collects is given by

$$R_n^*(c, p, r) = \max \left\{ rm : \ m \in \{0, 1, \ldots, n\} \text{ and } \sum_{\ell=1}^{m} W_{(\ell,n)} \leqslant c \right\}. \tag{2.1}$$

Here we compare the total reward of the offline-sort algorithm (2.1), $R_n^*(c, p, r)$, with that of an online feasible policy $\widehat{\pi} \in \Pi(n, c, p)$ that is based on a sequence of reoptimized time- and state-dependent threshold functions $\widehat{h}_n, \widehat{h}_{n-1}, \ldots, \widehat{h}_1$. If the current level of remaining capacity is x and the weight of item i is about to be revealed, then the decision maker computes the threshold $\widehat{h}_{n-i+1} : [0, \infty) \to [0, \infty)$ such that $\widehat{h}_{n-i+1}(x) \leqslant x$, and she selects item i if and only if the weight

$W_i \leqslant \widehat{h}_{n-i+1}(x)$. Thus if $\widehat{X}_0 = c$ and for $i \in [n]$ one defines the remaining capacity process $\widehat{X}_i$ recursively by

$$\widehat{X}_i = \begin{cases} \widehat{X}_{i-1} & \text{if } W_i > \widehat{h}_{n-i+1}(\widehat{X}_{i-1}) \\ \widehat{X}_{i-1} - W_i & \text{if } W_i \leqslant \widehat{h}_{n-i+1}(\widehat{X}_{i-1}), \end{cases}$$

then the total reward collected by the reoptimized policy $\widehat{\pi}$ can be written as

$$R_n^{\widehat{\pi}}(c,p,r) = \sum_{i=1}^{n} r\,\mathbb{1}\left\{W_i \leqslant \widehat{h}_{n-i+1}(\widehat{X}_{i-1})\right\}.$$

The random variables $R_n^*(c,p,r)$ and $R_n^{\widehat{\pi}}(c,p,r)$ crucially depend on the weight distribution F. This dependence is mostly expressed through a *consumption function* $\epsilon_{kp} : [0,\infty) \to [0,\infty]$ that is defined for $p \in (0,1]$ and for all $1 \leqslant k < \infty$ by

$$\epsilon_{kp}(x) = \sup\left\{\epsilon \in [0,\infty) : \int_0^{\epsilon} w\,dF(w) \leqslant \frac{x}{kp}\right\}. \tag{2.2}$$

The consumption function depends on two quantities. The argument x that denotes the current level of remaining capacity of the knapsack, and the index kp that refers to the expected number of items with F-distributed weights (or *arrivals*) that are yet to be presented to the decision maker. Furthermore, the function $\epsilon_{kp}(x)$ is always well defined. If $\mu = \mathbb{E}[\mathfrak{W}_1] = \mathbb{E}[W_1 \mid B_1 = 1]$ and $kp\mu < x < \infty$ then $\epsilon_{kp}(x) = +\infty$. Otherwise, the value $\epsilon_{kp}(x)$ satisfies the integral representation

$$\int_0^{\epsilon_{kp}(x)} w\,dF(w) = \frac{x}{kp} \qquad \text{for all } x \in [0, kp\mu]. \tag{2.3}$$

The representation (2.3) offers an important insight regarding the role of the consumption function $\epsilon_{kp}(x)$. The integral on the left-hand side is the *expected reduction* in the remaining capacity of the knapsack when the current level of remaining capacity is equal to x, and the decision maker selects an item with weight smaller than $\epsilon_{kp}(x)$. The function $\epsilon_{kp}(x)$ is then defined so that the expected reduction in capacity is equal to the ratio of the current capacity, x, to the expected number of remaining arrivals, kp. That is, the threshold $\epsilon_{kp}(x)$ is constructed so that—in expectation—the available capacity is spread equally over the remaining arrivals.

As we will see shortly, the threshold $\epsilon_{kp}(x)$ drives most of the estimates in this chapter and, together with the continuity of the weight distribution F, it immediately provides us with an easy upper bound for $\mathbb{E}[R_n^*(c,p,r)]$. The same threshold together with some mild regularity conditions

on the weight distribution F also drives the lower bound for $\mathbb{E}\left[R_n^{\hat{\pi}}(c,p,r)\right]$. The class of weight distributions we consider for the lower bound is characterized in the next definition.

Definition 2.1 (Typical class of distributions with continuous density). We say that a non-negative distribution F with continuous density function f belongs to the *typical class* if for some $\bar{w} > 0$, the following two conditions hold.

(i) Behavior at zero. There are $0 < \lambda < 1$ and $0 < \gamma < 1$ such that

$$\frac{F(\lambda w)}{F(w)} \leqslant \gamma < 1 \qquad \text{for all } w \in (0, \bar{w}). \tag{2.4}$$

(ii) Monotonicity. The map $w \mapsto w^3 f(w)$ is non-decreasing on $(0, \bar{w})$. That is,

$$w_1^3 f(w_1) \leqslant w_2^3 f(w_2) \qquad \text{for all } 0 < w_1 \leqslant w_2 < \bar{w}. \tag{2.5}$$

The class of typical distributions is wide enough to include most well-known non-negative distributions. In Section 2.4, we provide specific examples as well as class properties, but for now we emphasize that the breadth of the typical class comes from the role of the distribution-dependent parameter $\bar{w} > 0$. Conditions (2.4) and (2.5) need only to hold near zero—or, more precisely, on $(0, \bar{w})$—and not on the full support of the weight distribution or on the whole capacity interval $[0, c]$. In fact, for many distributions the parameter $\bar{w}$ for which (2.4) and (2.5) hold is much smaller than the minimum between the initial capacity and the supremum of the support.

The main results of this chapter are gathered in the theorem below. First, we provide an upper bound for $\mathbb{E}\left[R_n^*(c,p,r)\right]$ that holds for any continuous distribution F. Then, we turn to distributions that belong to the typical class, and we prove that there is a matching lower bound. As a by-product of our analysis, we establish that the regret is, at most, $O(\log n)$ as $n \to \infty$.[2] While our theoretical result provides only a regret bound, related results and the numerical experiments of Section 2.6 tell us that the regret bound is actually of the correct order.

Theorem 2.1 (Logarithmic regret bound)**.** *Consider a knapsack problem with capacity $0 \leqslant c < \infty$ and with items that arrive over $1 \leqslant n < \infty$ periods according to a Bernoulli process with arrival probability $p \in (0, 1]$. If the items have rewards equal to r and weights with continuous distribution F, then*

$$\max_{\pi \in \Pi(n,c,p)} \mathbb{E}\left[R_n^{\pi}(c,p,r)\right] \leqslant \mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant nprF(\epsilon_{np}(c)).$$

[2]Throughout this chapter, the function log denotes the natural logarithm.

Furthermore, there is a feasible online policy $\widehat{\pi} \in \Pi(n,c,p)$ *such that if the weights are independent and their distribution* F *belongs to the typical class then there is a constant* $1 < M < \infty$ *depending only on* F*,* p*, and* r *for which*

$$nprF(\epsilon_{np}(c)) - M(1+\log n) \leqslant \mathbb{E}\left[R_n^{\widehat{\pi}}(c,p,r)\right] \leqslant \max_{\pi\in\Pi(n,c,p)} \mathbb{E}\left[R_n^{\pi}(c,p,r)\right].$$

In turn, if the weights are independent and the distribution F *belongs to the typical class, then we have the regret bound*

$$\mathbb{E}\left[R_n^{*}(c,p,r)\right] - \max_{\pi\in\Pi(n,c,p)} \mathbb{E}\left[R_n^{\pi}(c,p,r)\right] \leqslant \mathbb{E}\left[R_n^{*}(c,p,r)\right] - \mathbb{E}\left[R_n^{\widehat{\pi}}(c,p,r)\right] \leqslant M(1+\log n).$$

The special case with deterministic arrivals and unitary rewards has been extensively studied in the literature. The upper bound $\mathbb{E}\left[R_n^{*}(c,1,1)\right] \leqslant nF(\epsilon_n(c))$ was first proved by Bruss and Robertson (1991). Here, we provide a generalization that is based on a relaxation of some appropriate optimization problem. The solution to this relaxation is the basis for constructing the reoptimized heuristics $\widehat{\pi}$. The lower bound $\mathbb{E}\left[R_n^{\widehat{\pi}}(c,p,r)\right] \geqslant nprF(\epsilon_{np}(c)) - O(\log n)$ as $n \to \infty$ is essentially new, and it substantially improves on existing estimates. The best results to date for general weight distribution F are due to Rhee and Talagrand (1991) who study a non-adaptive heuristic and prove that

$$nF(\epsilon_n(c))\left\{1 - \left[\frac{\epsilon_n(c)}{c}\right]^{1/2} - \frac{\epsilon_n(c)}{c}\right\} \leqslant \max_{\pi\in\Pi(n,c,1)} \mathbb{E}\left[R_n^{\pi}(c,1,1)\right] \qquad \text{for all } n \geqslant 1. \tag{2.6}$$

For instance, if $F(x) = \sqrt{x}$ for $x \in (0,1)$ then the lower bound (2.6) implies an upper bound for the regret that is $O(n^{1/3})$ as $n \to \infty$. Similarly, if $F(x) = x^2$ for $x \in (0,1)$ then the same lower bound gives us a regret upper bound that behaves like $O(n^{1/6})$ as $n \to \infty$.

A case that deserves special attention is when F is the uniform distribution on the unit interval, the reward $r = 1$, and the initial capacity $c = 1$. In this context, the Rhee and Talagrand (1991) lower bound provides us with a regret upper bound that behaves like $O(n^{1/4})$ as $n \to \infty$, but better bounds are available in the literature. This special dynamic and stochastic knapsack problem is in fact equivalent to the problem of the sequential selection of a monotone decreasing subsequence from a sample of n independent observation with the uniform distribution on the unit interval (cf. Samuels and Steele, 1981). The equivalence was first observed by Coffman et al. (1987, pp. 457–458), and it can be established by observing that the Bellman equations for the two problems are the

same after a change of variable. Informally, if the number of remaining periods is the same in both problems and the current capacity of the knapsack is equal to the last selected subsequence element, then the largest weight that is optimal for inclusion is equal to the maximum amount the decision maker is willing to go down in optimally selecting a new subsequence element. Since the weights as well as the subsequence elements are both uniformly distributed on the unit interval, these two actions happen with the same probability. For this subsequence-selection problem, Arlotto et al. (2015, 2018) prove that the expected performance ν_n^* of the best online policy satisfies the estimate $\nu_n^* = \sqrt{2n} - O(\log n)$ as $n \to \infty$. The equivalence between the two problems, however, holds *only* for uniform weights. As Theorem 2.1 suggests, the weight distribution F plays a crucial role in the estimates for the dynamic and stochastic knapsack problem with equal rewards. Instead, the monotone subsequence problem is distribution invariant, and one can consider uniformly distributed subsequence elements without loss of generality. More importantly, Seksenbayev (2018) and Gnedin and Seksenbayev (2019) characterize the second order asymptotic expansion of ν_n^* and establish that $\nu_n^* = \sqrt{2n} - \frac{1}{12}\log n + O(1)$ as $n \to \infty$. This remarkable result tells us that our regret bound is order tight, and that no online algorithm can—at this level of generality—be within $O(1)$ of offline sort.

Organization of the Chapter

The rest of this chapter is organized as follows. In Section 2.1, we review the related literature. In Section 2.2, we prove the prophet upper bound $\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant nprF(\epsilon_{np}(c))$ by showing that the offline-sort algorithm (2.1) can be reinterpreted as a parsimonious threshold policy and by solving a relaxation of some related optimization problem. This solution then guides us in the construction of policy $\widehat{\pi}$ that is presented in Section 2.3. In Section 2.4, we discuss the generality of the typical class of distributions, and we derive some properties that we then use—in Section 2.5—to prove that the reoptimized policy $\widehat{\pi}$ exhibits logarithmic regret. In Section 2.6, we present numerical experiments that provide further insights into our regret bound, while in Section 2.7 we discuss weight distributions with multiple types. Finally, in Section 2.8 we make closing remarks and underscore some open problems.

2.1 Literature Review: Knapsack Problems and Approximations

Knapsack problems uniquely combine simple formulations, non-trivial mathematical analyses, and relevance in several application-driven domains. As such, different knapsack problems have been considered in the literature, and a lot of effort has been devoted to the development of (near-) optimal policies. Most of the differences that have been accounted for concern the item arrival process (*static* versus *dynamic*), the probabilistic assumptions on the weight-reward pairs (deterministic and/or stochastic), and the objective of the decision maker (reward maximization, target achievement, etc.).

For instance, in the early formulation of Dantzig (1957), we have a *static* model with a finite number of items that are all available before any decision is made and have deterministic weights and deterministic rewards. The decision maker then seeks to find a maximum-reward subset of these items with total weight that does not exceed a capacity constraint. Following this classic formulation, researchers have considered several *static* knapsack instances with randomness in the weights and/or in the rewards. While studying a scheduling problem, Derman et al. (1978) studied a *static* and stochastic knapsack problem with items that belong to different categories. Items that belong to the same category have common deterministic rewards and independent, exponentially distributed weights with category-dependent parameter. The decision maker then seeks to maximize total expected rewards when the realized weights are revealed only after each item is included in the knapsack. The authors prove that the greedy policy based on reward-to-mean-weight ratios is optimal. Analogous *static* and stochastic knapsack problems have been considered by several authors, including Dean et al. (2004, 2005, 2008), Bhalgat et al. (2011), Li and Yuan (2013), Blado et al. (2016), Ma (2018), Blado and Toriello (2019), and Balseiro and Brown (2019). Gupta et al. (2011) and Merzifonluoğlu et al. (2012) follow along similar lines, but consider both random weights and random rewards. Most notably, Dean et al. (2004, 2005, 2008) study a *static* and stochastic knapsack problem with deterministic rewards and independent random weights with arbitrary distributions that are realized only upon insertion in the knapsack. They construct a polynomial time adaptive policy that is within a constant multiplicative gap, and they compare the performance of adaptive and non-adaptive policies. Their work is particularly relevant to us as it is among the first ones to assess the benefits of adaptivity.

Static stochastic knapsack problems have also been studied under different optimization objectives. For instance, there is a stream of related literature that considers *static* stochastic knapsack problems (typically with deterministic weights and random rewards) in which the objective is to maximize the probability that the total reward will achieve a certain given target. (See, e.g., Henig, 1990; Carraway et al., 1993; Ilhan et al., 2011, among others.)

Alongside the static knapsack problems mentioned thus far there are several *dynamic* models in which items arrive over time and their weight-reward pairs are revealed to the decision maker who irrevocably decides on inclusion in the knapsack as soon as each item arrives and without seeing the weights and/or the rewards of future items. *Dynamic* and stochastic knapsack problems are widespread. For instance, if one assumes that the weights are all equal to one and that the rewards are random, then one recovers the multi-secretary problem (see, e.g. Cayley, 1875; Moser, 1956; Kleinberg, 2005). For this problem, Arlotto and Gurvich (2019) prove that if the reward distribution is discrete, then the regret is uniformly bounded in the number of items and the knapsack capacity. Similarly, if one assumes that the rewards are all equal to one and that the weights are random, then one finds an instance of the single-machine scheduling problem of Baruah et al. (1994) that motivates this chapter. Finally, when both the weights and the rewards are random, one recovers—among others—the sequential investment problems of Derman et al. (1975) and Prastacos (1983), or the multi-secretary problem of Nakai (1986) which allows for an unknown number of applicants in each period. When both the weights and the rewards are random, few regret bounds are available. A notable exception is the work of Marchetti-Spaccamela and Vercellis (1995) who prove a $O(\log^{3/2} n)$ regret bound when both the weights and the rewards are independent and uniformly distributed on the unit interval, and the knapsack capacity is proportional to the number of periods. For the same formulation, Lueker (1998) improves Marchetti-Spaccamela and Vercellis's result to $O(\log n)$ and shows that it is best possible.

Multi-dimensional generalizations of the *dynamic* and stochastic knapsack problem have found several applications in revenue management and resource allocation. In the network revenue management problem, heterogeneous customers belonging to different classes arrive sequentially over time, request a product, and offer a price. If the request is accepted, then a collection of resources that constitute the product is depleted, and the offered price is earned. Otherwise the resource capacities remain unchanged and the offered price is lost (cf. Gallego and van Ryzin, 1997; Talluri and van Ryzin, 2004). The solution of the network revenue management problem is famously diffi-

cult, and scholars have studied several non-adaptive as well as adaptive heuristics and proved regret bounds. A classic non-adaptive approximation scheme based on a deterministic linear-programming relaxation was studied by Gallego and van Ryzin (1994, 1997). In contrast, adaptive policies have been considered by allowing for periodic reoptimization. Despite a few specific negative results by Cooper (2002), Chen and Homem-de Mello (2010), and Jasin and Kumar (2013), there are ways to construct reoptimized policies that perform well. For instance, Reiman and Wang (2008) propose a probabilistic allocation rule that works well with one reoptimization instance. Jasin and Kumar (2012) and Wu et al. (2015) consider a probabilistic allocation rule that is based on reoptimizing in every period and show that it exhibits uniformly bounded regret provided that the optimal solution to the original deterministic linear programming relaxation is non-degenerate. Bumpensanti and Wang (2020) and Vera and Banerjee (2019) prove that the uniform regret bound holds in general, without the non-degeneracy assumption.

2.2 A Prophet Upper Bound

The performance of any online algorithm is bounded above by the full-information (or offline) sort. If the decision maker knows all of the weights $W_1, W_2, \ldots, W_n$ before making any decision, then the total reward she collects is the largest number rm such that the sum of the smallest m realizations does not exceed the capacity constraint. That is, if $W_{(1,n)} \leqslant W_{(2,n)} \leqslant \cdots \leqslant W_{(n,n)}$ are the order statistics of $\mathcal{W} \equiv \{W_1, W_2, \ldots, W_n\}$, then the total reward $R_n^*(c, p, r)$ of offline selections when the initial knapsack capacity is c and the arrival probability is p is given by

$$R_n^*(c,p,r) = \max\left\{ rm : \ m \in \{0,1,\ldots,n\}, \sum_{\ell=1}^{m} W_{(\ell,n)} \leqslant c \text{ and } W_{(\ell,n)} \in \mathcal{W} \text{ for all } \ell \in [n] \right\}. \quad (2.7)$$

Earlier work has considered unitary rewards and deterministic arrivals by studying the random variable $R_n^*(c, 1, 1)$. First along this line of research, Coffman et al. (1987) showed that

$$R_n^*(c,1,1) \sim nF(\epsilon_n(c)) \text{ in probability as } n \to \infty,$$

provided that the weight distribution F is continuous, strictly increasing in w when $F(w) < 1$, and $F(w) \sim Aw^\alpha$ as $w \to 0$ for some $A, \alpha > 0$. Four years later, Bruss and Robertson (1991) proved that the same result holds under more general conditions, and Boshuizen and Kertz (1999) established the asymptotic normality of $R_n^*(c, 1, 1)$ after the usual centering and scaling for different classes of

weight distribution F. Lemma 4.1 in Bruss and Robertson (1991) is particularly relevant to our discussion here since it tells us that

$$\mathbb{E}\left[R_n^*(c,1,1)\right] \leqslant nF(\epsilon_n(c)) \qquad \text{for all } n \geqslant 1.$$

Here, we generalize this result by accounting for Bernoulli arrivals with probability $p \in (0,1]$ and rewards equal to $r > 0$. Specifically, we show that

$$\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant nprF(\epsilon_{np}(c)) \qquad \text{for all } n \geqslant 1.$$

Our proof relies on the observation that the offline-sort algorithm (2.7) can be equivalently described as an algorithm that selects items with weight that is below some threshold. For any given realization $W_1, W_2, \ldots, W_n$, the offline-sort algorithm selects $N_n^* \equiv R_n^*(c,p,1)$ items so one can compute the value $W_{(N_n^*,n)}$ of the largest weight that is selected for inclusion, and one can then select all of the items $i \in [n]$ that have weight $W_i \leqslant W_{(N_n^*,n)}$. A shortcoming of this interpretation is that one needs to know the realization of the weight W_i (as well as the realizations of all of the other weights) to compute the threshold $W_{(N_n^*,n)}$. As it turns out, this is not needed in general. The next lemma shows that there is a thresholding algorithm that makes the same selections of offline sort, but in which the threshold used to decide whether to select an item is computed without using the information about that item's weight.

Lemma 2.1 (Threshold policy equivalence)**.** *Let $W_{(1,n)} \leqslant W_{(2,n)} \leqslant \cdots \leqslant W_{(n,n)}$ be the order statistics of $\mathcal{W} \equiv \{W_1, W_2, \ldots, W_n\}$ and, for $i \in [n]$, let $W_{(1,n-1)} \leqslant W_{(2,n-1)} \leqslant \cdots \leqslant W_{(n-1,n-1)}$ be the order statistics of $\mathcal{W}_i = \mathcal{W} \backslash \{W_i\}$. Then, for*

$$\tau_{n-1}^i = \max\left\{ m \in \{0,1,\ldots,n-1\} : \ \sum_{\ell=1}^{m} W_{(\ell,n-1)} \leqslant c \text{ and } W_{(\ell,n-1)} \in \mathcal{W}_i \text{ for all } \ell \in [n-1] \right\} \tag{2.8}$$

and $N_n^* \equiv R_n^*(c,p,1)$, we have that

$$W_i \leqslant W_{(N_n^*,n)} \qquad \text{if and only if} \qquad W_i \leqslant h(\mathcal{W}_i) \equiv \max\left\{W_{(\tau_{n-1}^i,n-1)}, c - \sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n-1)}\right\}. \tag{2.9}$$

In turn, it follows that

$$R_n^*(c,p,r) = \sum_{i=1}^{n} r\mathbb{1}\left\{W_i \leqslant h(\mathcal{W}_i)\right\}. \tag{2.10}$$

Proof. The equivalence (2.10) is an obvious consequence of (2.9), so we focus on proving the latter. If $N_n^* = n$ we have that $\tau_{n-1}^i = n-1$ and $W_i \leqslant c - \sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n-1)}$ for all $i \in [n]$, so equivalence (2.9) immediately follows. Instead, if $N_n^* < n$ the proof of (2.9) requires more work. As a warm-up we note that since the sets $\mathcal{W}$ and $\mathcal{W}_i$ differ only in one element, then

$$W_{(\ell,n)} \leqslant W_{(\ell,n-1)} \leqslant W_{(\ell+1,n)} \qquad \text{for all } \ell \in [n-1]. \tag{2.11}$$

If we now recall the definitions of τ_{n-1}^i and N_n^* and use the inequalities above we obtain that

$$\sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n)} \leqslant \sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n-1)} \leqslant c \qquad \text{and} \qquad \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n-1)} \leqslant \sum_{\ell=1}^{N_n^*-1} W_{(\ell+1,n)} \leqslant \sum_{\ell=1}^{N_n^*} W_{(\ell,n)} \leqslant c.$$

These two bounds respectively tell us that the offline-sort algorithm on $\mathcal{W}$ selects at least τ_{n-1}^i observations, and that the same algorithm on $\mathcal{W}_i$ selects at least $N_n^* - 1$ items. Thus, it follows that

$$N_n^* - 1 \leqslant \tau_{n-1}^i \leqslant N_n^*,$$

and we use these bounds to prove the equivalence (2.9).

If. We now suppose that $W_i \leqslant h(\mathcal{W}_i) \equiv \max\left\{W_{(\tau_{n-1}^i,n-1)}, c - \sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n-1)}\right\}$, and we seek to show that $W_i \leqslant W_{(N_n^*,n)}$. We consider two cases, one per each possible realization of τ_{n-1}^i.

Case 1: $\tau_{n-1}^i = N_n^* - 1$. If $\tau_{n-1}^i = N_n^* - 1$ then the definition of τ_{n-1}^i in (2.8) tells us that

$$c - \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n-1)} < W_{(N_n^*,n-1)},$$

so if we apply the right inequality of (2.11) to $\ell = N_n^* - 1$ and $\ell = N_n^*$, we obtain that

$$W_{(N_n^*-1,n-1)} \leqslant W_{(N_n^*,n)} \qquad \text{and} \qquad c - \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n-1)} < W_{(N_n^*+1,n)}. \tag{2.12}$$

If $W_{(N_n^*,n)} = W_{(N_n^*+1,n)}$ then the two inequalities in (2.12) give us that

$$h(\mathcal{W}_i) = \max\left\{W_{(N_n^*-1,n-1)}, c - \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n-1)}\right\} \leqslant W_{(N_n^*,n)},$$

so we also have that $W_i \leqslant W_{(N_n^*,n)}$. On the other hand, if $W_{(N_n^*,n)} < W_{(N_n^*+1,n)}$ then the bounds in (2.12) imply that $h(\mathcal{W}_i) < W_{(N_n^*+1,n)}$, so we obtain from $W_i \leqslant h(\mathcal{W}_i)$ that $W_i \leqslant W_{(N_n^*,n)}$.

Case 2: $\tau_{n-1}^i = N_n^*$**.** The left inequality of (2.11) with $\ell = N_n^*$ tells us that we have two sub-cases to consider here: (i) when $W_{(N_n^*,n)}$ is equal to $W_{(N_n^*,n-1)}$, and (ii) when $W_{(N_n^*,n)}$ is strictly smaller than $W_{(N_n^*,n-1)}$. In the first sub-case, if $\tau_{n-1}^i = N_n^*$ and $W_{(N_n^*,n)} = W_{(N_n^*,n-1)}$, then the first N_n^* order statistics of $\mathcal{W}$ and of $\mathcal{W}_i$ agree and $c - \sum_{\ell=1}^{N_n^*} W_{(\ell,n-1)} = c - \sum_{\ell=1}^{N_n^*} W_{(\ell,n)} < W_{(N_n^*+1,n)}$. Thus, if $W_{(N_n^*,n)} = W_{(N_n^*+1,n)}$ then $h(\mathcal{W}_i) = \max\{W_{(N_n^*,n)}, c - \sum_{\ell=1}^{N_n^*} W_{(\ell,n)}\} = W_{(N_n^*,n)}$, and we are done. Otherwise, if $W_{(N_n^*,n)} < W_{(N_n^*+1,n)}$ then $h(\mathcal{W}_i) < W_{(N_n^*+1,n)}$ so that $W_i \leqslant h(\mathcal{W}_i) < W_{(N_n^*+1,n)}$ implies that $W_i \leqslant W_{(N_n^*,n)}$. In the second sub-case, if $\tau_{n-1}^i = N_n^*$ and $W_{(N_n^*,n)} < W_{(N_n^*,n-1)}$ then we have that $W_i = W_{(N_n^*,n)}$, and the result follows.

Only If. We now suppose that $W_i \leqslant W_{(N_n^*,n)}$, and we show that

$$W_i \leqslant h(\mathcal{W}_i) \equiv \max\left\{W_{(\tau_{n-1}^i,n-1)}, c - \sum_{\ell=1}^{\tau_{n-1}^i} W_{(\ell,n-1)}\right\}$$

by proving that $W_{(N_n^*,n)} \leqslant h(\mathcal{W}_i)$. Just as before, we consider separately the two possible realizations of τ_{n-1}^i.

Case 1: $\tau_{n-1}^i = N_n^* - 1$**.** We have two sub-cases to consider here. First, if $W_{(N_n^*,n)} \leqslant W_{(N_n^*-1,n-1)}$ then the lower bound $W_{(N_n^*,n)} \leqslant h(\mathcal{W}_i)$ is trivial. Second, if $W_{(N_n^*-1,n-1)} < W_{(N_n^*,n)}$ we show that the right maximand is bounded below by $W_{(N_n^*,n)}$. In this instance, the first $N_n^* - 1$ order statistic of $\mathcal{W}$ and $\mathcal{W}_i$ agree so the definition of N_n^* gives us that $W_{(N_n^*,n)} \leqslant c - \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n)} = c - \sum_{\ell=1}^{N_n^*-1} W_{(\ell,n-1)}$, and we are done.

Case 2: $\tau_{n-1}^i = N_n^*$**.** If $\tau_{n-1}^i = N_n^*$ the left inequality of (2.11) tells us that $W_{(N_n^*,n)} \leqslant W_{(N_n^*,n-1)}$, so the lower bound $W_{(N_n^*,n)} \leqslant h(\mathcal{W}_i)$ immediately follows.

□

The representation (2.10) for $R_n^*(c,p,r)$ provides us with an easy way for proving the upper bound $\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant nprF(\epsilon_{np}(c))$. We just need to note that the expected total reward collected by the offline-sort algorithm is bounded above by the solution of some appropriate optimization problem. Our argument does not require independence of item weights. The threshold equivalence of Lemma 2.1 holds on every sample path, and the relaxation that follows only uses properties of the weight distribution F and of the arrival probability p (see also Steele, 2016).

Proposition 2.1 (Prophet upper bound)**.** *Consider a knapsack problem with capacity* $0 \leqslant c < \infty$ *and with items that arrive over* $1 \leqslant n < \infty$ *periods according to a Bernoulli process with arrival probability* $p \in (0, 1]$*. If the items have rewards equal to* r *and weights with continuous distribution* F*, then for* $\epsilon_{np}(c) = \sup\left\{\epsilon \in [0, \infty) : \int_0^\epsilon w\, dF(w) \leqslant \frac{c}{np}\right\}$ *we have that*

$$\mathbb{E}\left[R_n^*(c, p, r)\right] \leqslant nprF(\epsilon_{np}(c)). \tag{2.13}$$

Proof. To prove inequality (2.13), we begin with two easy cases. If $c = 0$ then $R_n^*(0, p, r) = 0$, and the bound (2.13) is trivial. Similarly, if $\mu = \mathbb{E}\left[W_1 \mid B_1 = 1\right] = \int_0^\infty w\, dF(w)$ and $np\mu < c < \infty$ then the definition of the function $\epsilon_{np}(c)$ tells us that $\epsilon_{np}(c) = +\infty$ so $F(\epsilon_{np}(c)) = 1$ and the bound (2.13) is again trivial because $R_n^*(c, p, r) \leqslant \sum_{i=1}^n r\mathbb{1}\left\{W_i < \infty\right\}$ for all $c \in [0, \infty)$, and this last right-hand side has expected value equal to npr.

Next, we consider the case in which $0 < c \leqslant np\mu$. If $\mathcal{W}_i \equiv \{W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_n\}$ and $\mathcal{G}_i = \sigma\{\mathcal{W}_i\}$ is the σ-field generated by the sample $\mathcal{W}_i$, then we obtain from Lemma 2.1 and from the definition (2.7) that for each $i \in [n]$ there is a $\mathcal{G}_i$-measurable threshold $h(\mathcal{W}_i)$ such that one has the representation as well as the capacity constraint

$$R_n^*(c, p, r) = \sum_{i=1}^n r\mathbb{1}\left\{W_i \leqslant h(\mathcal{W}_i)\right\} \qquad \text{and} \qquad \sum_{i=1}^n W_i\mathbb{1}\left\{W_i \leqslant h(\mathcal{W}_i)\right\} \leqslant c.$$

In turn, we can obtain an upper bound for $\mathbb{E}\left[R_n^*(c, p, r)\right]$ by maximizing the sum $\sum_{i=1}^n \mathbb{E}\left[r\mathbb{1}\left\{W_i \leqslant h_i\right\}\right]$ over all thresholds $(h_1, h_2, \ldots, h_n)$ that satisfy an analogous capacity constraint and that have the same measurability property. Formally, we have the inequality

$$\begin{aligned} \mathbb{E}\left[R_n^*(c, p, r)\right] \leqslant \quad & \max_{(h_1, \ldots, h_n)} \quad \sum_{i=1}^n \mathbb{E}\left[r\mathbb{1}\left\{W_i \leqslant h_i\right\}\right] \\ & \text{s.t.} \qquad \sum_{i=1}^n W_i\mathbb{1}\left\{W_i \leqslant h_i\right\} \leqslant c \quad \text{almost surely} \\ & \qquad\quad h_i \in \mathcal{G}_i \quad \text{for all } i \in [n]. \end{aligned} \tag{2.14}$$

Since $\epsilon_{np}(c) > 0$ and because the capacity constraint holds almost surely (and thus also in expecta-

tion), we have the further upper bound

$$\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant \max_{(h_1,\ldots,h_n)} \sum_{i=1}^{n} \mathbb{E}\left[r\mathbb{1}\left\{W_i \leqslant h_i\right\}\left\{1-\epsilon_{np}^{-1}(c)W_i\right\}\right] + cr\epsilon_{np}^{-1}(c) \tag{2.15}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} W_i \mathbb{1}\left\{W_i \leqslant h_i\right\} \leqslant c \quad \text{almost surely}$$

$$h_i \in \mathcal{G}_i \quad \text{for all } i \in [n].$$

Because h_i is $\mathcal{G}_i$-measurable, an application of the tower property gives us that

$$\mathbb{E}\left[\mathbb{E}\left[r\mathbb{1}\left\{W_i \leqslant h_i\right\}\left\{1-\epsilon_{np}^{-1}(c)W_i\right\} \mid \mathcal{G}_i\right]\right] = pr\mathbb{E}\left[\int_0^{h_i}\left\{1-\epsilon_{np}^{-1}(c)w\right\}dF(w)\right],$$

so, after we drop the two constraints in (2.15) we obtain that

$$\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant \mathfrak{p}^* = \max_{(h_1,\ldots,h_n)} \sum_{i=1}^{n} pr\mathbb{E}\left[\int_0^{h_i}\left\{1-\epsilon_{np}^{-1}(c)w\right\}dF(w)\right] + cr\epsilon_{np}^{-1}(c). \tag{2.16}$$

The maximization problem on the right-hand side is separable, and it is simple task to verify that the quantity $\mathbb{E}\left[\int_0^{h_i}\left\{1-\epsilon_{np}^{-1}(c)w\right\}dF(w)\right]$ is maximized by setting $h_i = \epsilon_{np}(c)$ almost surely and for all $i \in [n]$. Thus, it follows that

$$\begin{aligned}\mathfrak{p}^* &= \sum_{i=1}^{n} \max_{h_i} pr\mathbb{E}\left[\int_0^{h_i}\left\{1-\epsilon_{np}^{-1}(c)w\right\}dF(w)\right] + cr\epsilon_{np}^{-1}(c) \\ &= npr\left\{F(\epsilon_{np}(c)) - \epsilon_{np}^{-1}(c)\left[\int_0^{\epsilon_{np}(c)} w\,dF(w) - \frac{c}{np}\right]\right\}.\end{aligned}$$

The integral representation (2.3) then tells us that the second summand is equal to zero, so after we recall (2.16) we obtain that

$$\mathbb{E}\left[R_n^*(c,p,r)\right] \leqslant \mathfrak{p}^* = nprF(\epsilon_{np}(c)) \qquad \text{for all } 0 < c \leqslant np\mu,$$

completing the proof of (2.13). □

2.3 The Reoptimized Policy $\widehat{\pi}$ and Its Value Function

In the course of proving Proposition 2.1, we observed that if $\mathcal{G}_i = \sigma\{W_1,\ldots,W_{i-1},W_{i+1},\ldots,W_k\}$ is the σ-field generated by the sample $\{W_1,\ldots,W_{i-1},W_{i+1},\ldots,W_k\}$, then the expected value of the

offline solution $R_k^*(x,p,r)$ satisfies the upper bound

$$\begin{aligned}\mathbb{E}\left[R_k^*(x,p,r)\right] \leqslant \quad & \max_{(h_1,\ldots,h_k)} \quad \sum_{i=1}^{k} \mathbb{E}\left[r\,\mathbb{1}\left\{W_i \leqslant h_i\right\}\right] \\ & \text{s.t.} \qquad \sum_{i=1}^{k} W_i\,\mathbb{1}\left\{W_i \leqslant h_i\right\} \leqslant x \quad \text{almost surely} \\ & \qquad\quad h_i \in \mathcal{G}_i \quad \text{for all } i \in [k].\end{aligned}$$

We also noticed that the optimization problem on the right-hand side can be relaxed by first adding to its objective the quantity $\epsilon_{kp}^{-1}(x)r\left\{x - \mathbb{E}\left[\sum_{i=1}^{k} W_i\,\mathbb{1}\left\{W_i \leqslant h_i\right\}\right]\right\} \geqslant 0$, and then by dropping the two constraints. This then gives us the further upper bound

$$\mathbb{E}\left[R_k^*(x,p,r)\right] \leqslant \max_{(h_1,\ldots,h_k)} \sum_{i=1}^{k} pr\mathbb{E}\left[\int_0^{h_i} \left\{1 - \epsilon_{kp}^{-1}(c)w\right\} dF(w)\right] + xr\epsilon_{kp}^{-1}(x), \tag{2.17}$$

which is maximized by setting $h_i = \epsilon_{kp}(x)$ for all $i \in [k]$. We can now use this reoptimized solution for all $x \in [0,\infty)$ and all $1 \leqslant k < \infty$ to construct the online feasible threshold policy $\widehat{\pi} \in \Pi(n,c,p)$. Specifically, since $\epsilon_{kp}(x)$ may exceed x, we set for $p \in (0,1]$

$$\widehat{h}_k(x) = \min\{x, \epsilon_{kp}(x)\}, \tag{2.18}$$

and we define the reoptimized policy $\widehat{\pi}$ through the threshold $\{\widehat{h}_n, \widehat{h}_{n-1}, \ldots, \widehat{h}_1\}$. Thus, if the remaining capacity is x when item i is first presented, then item i is selected if and only if its weight $W_i \leqslant \widehat{h}_{n-i+1}(x)$.

In turn, the threshold functions $\{\widehat{h}_k : 1 \leqslant k < \infty\}$ induce a sequence of value functions $\{\widehat{v}_k : [0,\infty) \to \mathbb{R}_+ : 0 \leqslant k < \infty\}$ such that $\widehat{v}_k(x)$ represents the expected reward to-go of the reoptimized policy when there are k remaining periods and the current level of remaining knapsack capacity is x. If $\widehat{v}_0(x) = 0$ for all $x \in [0,\infty)$, then the value $\widehat{v}_k(x)$ is given by the recursion

$$\begin{aligned}\widehat{v}_k(x) &= p\left(1 - F(\widehat{h}_k(x))\right)\widehat{v}_{k-1}(x) + p\int_0^{\widehat{h}_k(x)} \left\{r + \widehat{v}_{k-1}(x-w)\right\} dF(w) + (1-p)\widehat{v}_{k-1}(x) \\ &= \left(1 - pF(\widehat{h}_k(x))\right)\widehat{v}_{k-1}(x) + p\int_0^{\widehat{h}_k(x)} \left\{r + \widehat{v}_{k-1}(x-w)\right\} dF(w).\end{aligned} \tag{2.19}$$

By setting the number of remaining periods to n and the knapsack capacity to c, we find that

$$\widehat{v}_n(c) = \mathbb{E}\left[R_n^{\widehat{\pi}}(c,p,r)\right].$$

To verify the validity of the recursion (2.19), we condition on what happens in the kth-to-last period. With probability $1-p$ the arriving item has arbitrarily large weight (equivalently, no item arrives), the number of the remaining periods decreases to $k-1$ and the level of remaining capacity, x, stays the same. This then yields the term $(1-p)\widehat{v}_{k-1}(x)$ in the first line of (2.19). On the other hand, with probability p the arriving item has weight distribution F, and we can further condition on its realization, w. If $w > \widehat{h}_k(x)$ then the item is rejected, the level of remaining capacity does not change, and the number of remaining periods decreases by one. That is, if the item is rejected, the expected reward to-go is given by $\widehat{v}_{k-1}(x)$ and, since rejections happen with probability $p(1-F(\widehat{h}_k(x)))$, we recover the first summand on the top line of (2.19). On the other hand, if $w \leqslant \widehat{h}_k(x)$ the kth-to-last item is included in the knapsack. Such a decision produces an immediate reward of r, and it depletes w units of capacity. The new remaining capacity then becomes $x-w$, and the number of remaining periods decreases to $k-1$. The decision maker's payoff for including this item is then given by $r+\widehat{v}_{k-1}(x-w)$ and, by integrating this payoff against the measure $p\,dF(w)$ for $w \in [0, \widehat{h}_k(x)]$, we find the second summand on the first line of the recursion (2.19).

The reoptimized heuristic $\widehat{\pi}$ then takes the solution of the offline relaxation (2.17) and turns it into an online algorithm through the threshold $\widehat{h}_k$ given in (2.18). This direct link provides us with enough tractability to be able to quantify the difference in expected performance between the reoptimized heuristic and the offline solution and—as a result—to prove the logarithmic regret bound. Instead, the optimal dynamic programming policy cannot be expressed explicitly and it lacks of the regularity needed to make any meaningful analytical progress. However, we note here that both the reoptimized heuristic and the optimal dynamic programming policy can be computed numerically in polynomial time, and we refer the reader to Section 2.6 for more details on our numerical work.

2.4 On the Typical Class

The weight distribution F plays a crucial role in the study of the performance of optimal and near-optimal item selections for the dynamic and stochastic knapsack problem with equal rewards. Because the weights are not equal, the remaining capacity process exhibits substantial randomness, and this may lead to unexpected behavior. As such, regularity conditions on the weight distribution

F are commonplace in the related literature. For instance, Coffman et al. (1987) only consider distributions F such that $F(w) \sim Aw^{\alpha}$ as $w \to 0$ for some $A, \alpha > 0$, while Bruss and Robertson (1991) expand this class to include all of the weight distributions F such that $\limsup_{w \to 0^+} F(\lambda w)/F(w) < 1$. Furthermore, Papastavrou et al. (1996, Section 5) show that one must require concavity of F to obtain structural properties such as monotonicity of the optimal threshold functions and concavity of the optimal value functions.

Here, we consider distributions that belong to the typical class characterized in Definition 2.1. As we mentioned earlier, this class is broad enough to include most well-known non-negative continuous distributions. Such breadth comes from the fact that Conditions (2.4) and (2.5) in Definition 2.1 must hold only on $(0, \bar{w})$ for some $\bar{w} > 0$, and that one has the flexibility of choosing different parameter $\bar{w}$ for different distribution F. For instance, the uniform distribution $f(w) = \mathbb{1}\{w \in (0,1)\}$ and the exponential distribution $f(w) = \alpha e^{-\alpha w}\mathbb{1}\{w > 0\}$ are both typical, but they require different choices of $\bar{w}$. For the uniform distribution, Conditions (2.4) and (2.5) hold on all of its support and one can choose $\bar{w} = 1$, while for the exponential distribution, Condition (2.5) holds only on $(0, 3/\alpha)$ and one can set $\bar{w} = 3/\alpha$. Similarly, one can check that the truncated normal distribution on $(0, b)$ with density $f(w) = A\exp\{-(w-\upsilon)^2/(2\varsigma^2)\}\mathbb{1}\{w \in (0,b)\}$ for $\upsilon \in \mathbb{R}$, $\varsigma > 0$, and A being the appropriate normalizing constant, is typical with $\bar{w} = \min\{\frac{1}{2}(\upsilon + \sqrt{\upsilon^2 + 12\varsigma^2}), b\}$. The truncated logistic distribution on $(0, b)$ and the logit-normal distribution are additional examples of typical distributions, though the respective $\bar{w}$'s have to do with the smallest positive root of related transcendental equations. The families of distributions listed below also belong to the typical class.

1. *Power distributions.* Distributions such that $F(w) = Aw^{\alpha}$ for some $A, \alpha > 0$ on $(0, \bar{w})$ are typical. Condition (2.4) is immediately verified. The function $w^3 f(w) = A\alpha w^{\alpha+2}$ is increasing because $A, \alpha > 0$, so (2.5) holds as well.

2. *Convex distributions.* Distributions F that are convex in a neighborhood of 0 and that have continuous density f are typical. Convexity tells us that $F(\lambda w) \leqslant F(w)\lambda$ so (2.4) follows. Furthermore, convexity also gives us that the density f is non-decreasing, so (2.5) is verified.

3. *Mixtures of typical distributions.* The class of typical distributions is closed under mixture. If F and G are two typical distributions and $\beta \in [0,1]$ then it is easy to see that the mixture distribution $\beta F + (1-\beta)G$ is also typical.

It is important to note, however, that one can construct examples of distributions that do not belong to the typical class. For instance, the distribution $F(w) = \frac{\log \bar{w}}{\log w}$ for $\bar{w} < 1$ and $w \in (0, \bar{w})$ is an example that satisfies Condition (2.5) but violates Condition (2.4). For a fixed $0 < \lambda < 1$, one can easily check that

$$\limsup_{w \to 0^+} \frac{F(\lambda w)}{F(w)} = \limsup_{w \to 0^+} \frac{\log w}{\log \lambda + \log w} = 1,$$

so Condition (2.4) fails to hold. On the other hand, the function $w^3 f(w) = -\frac{w^2 \log \bar{w}}{(\log w)^2}$ is increasing on $(0, \bar{w})$ and Condition (2.5) is satisfied.

The distribution $F(w) = A \int_0^w \{\sin(1/u)\}^2 \, du$ for $w \in (0, \bar{w})$ and $A = (\int_0^{\bar{w}} \{\sin(1/u)\}^2 \, du)^{-1} > 0$ is an example that satisfies Condition (2.4) while violating Condition (2.5). In fact, one has that the limit

$$\limsup_{w \to 0^+} \frac{F(\lambda w)}{F(w)} = \lambda < 1,$$

but the function $w^3 f(w) = A w^3 \{\sin(1/w)\}^2$ oscillates infinitely many times in a (positive) neighborhood of zero, so the monotonicity (2.5) fails to hold.

We conclude this section by observing that Condition (2.4) regarding the behavior of F at zero is equivalent to the condition required by Bruss and Robertson (1991), and by proving that we can equivalently state it as a property of the ratio $wF(w)/\int_0^w u \, dF(u)$. This equivalent property will be important to our analysis.

Lemma 2.2 (Equivalence of CDF Conditions)**.** *There are constants $0 < \lambda < 1$ and $0 < \gamma < 1$ and a value $\bar{w} > 0$ such that*

$$\frac{F(\lambda w)}{F(w)} \leqslant \gamma < 1 \qquad \text{for all } w \in (0, \bar{w}) \tag{2.20}$$

if and only if there is a constant $1 < M < \infty$ such that

$$\frac{wF(w)}{\int_0^w u \, dF(u)} \leqslant M < \infty \qquad \text{for all } w \in (0, \bar{w}). \tag{2.21}$$

Proof. If. Suppose there is a constant $1 < M < \infty$ such that condition (2.21) holds. Next, note that for any $\lambda \in (0, 1)$ and any $w \in (0, \bar{w})$ one has the bounds

$$0 \leqslant \int_0^w u \, dF(u) \leqslant \lambda w \int_0^{\lambda w} dF(u) + w \int_{\lambda w}^w dF(u) = wF(w) - wF(\lambda w)(1 - \lambda),$$

so it follows that

$$\frac{F(w)}{F(w) - F(\lambda w)(1-\lambda)} \leqslant \frac{wF(w)}{\int_0^w u\, dF(u)}.$$

In turn, condition (2.21) tells us that there is $1 < M < \infty$ such that the right-hand side above is bounded by M so, after rearranging, we obtain that

$$\frac{F(\lambda w)}{F(w)} \leqslant \frac{M-1}{M(1-\lambda)} \qquad \text{for all } w \in (0, \bar{w}).$$

Condition (2.20) then follows after one chooses any $\lambda < M^{-1}$ and sets $\gamma = (M-1)/[M(1-\lambda)] < 1$.

Only if. Suppose that there are constants $0 < \lambda < 1$ and $0 < \gamma < 1$ such that condition (2.20) holds for some $\bar{w} > 0$. Then we have that

$$0 < 1 - \gamma \leqslant 1 - \frac{F(\lambda w)}{F(w)} = \int_{\lambda w}^{w} \frac{dF(u)}{F(w)} \qquad \text{for all } w \in (0, \bar{w}).$$

Moreover, if we multiply both sides by λw and use the fact that $\lambda w \leqslant u$ for all $u \in (\lambda w, w)$ we also have that

$$\lambda w(1-\gamma) \leqslant \int_{\lambda w}^{w} \frac{\lambda w}{F(w)}\, dF(u) \leqslant \int_0^w \frac{u}{F(w)}\, dF(u).$$

Next, we divide both sides by w and rearrange to obtain that

$$\frac{wF(w)}{\int_0^w u\, dF(u)} \leqslant \frac{1}{\lambda(1-\gamma)} \qquad \text{for all } w \in (0, \bar{w}),$$

so condition (2.21) follows by setting $M = [\lambda(1-\gamma)]^{-1}$, and the proof is now complete. □

2.5 A Logarithmic Regret Bound

To prove that the regret grows at most logarithmically, we let

$$K = \left\lceil \frac{c}{p \int_0^{\bar{w}} w f(w)\, dw} \right\rceil \tag{2.22}$$

and focus on dynamic and stochastic knapsack problems with more than K periods. Of course, this is without loss of generality because the quantity K defined in (2.22) is a constant that does not depend on the number of periods n, so we can ignore the last K decisions without affecting our

regret bound. When $k \geqslant K$ we have (i) that $\epsilon_{kp}(x) \leqslant \bar{w}$ for all $x \in [0, c]$, and (ii) that the integral representation (2.3) always holds. Thus, we are focusing on problem instances in which we can use the properties of the typical class in full.

In our proof, we will repeatedly use the following two properties of the consumption function $\epsilon_{kp}(x)$. First, we obtain from definition (2.2) that the consumption functions are non-increasing in k. That is, for $p \in (0, 1]$ one has the monotonicity

$$\epsilon_{(k+1)p}(x) \leqslant \epsilon_{kp}(x) \qquad \text{for all } x \in [0, \infty) \text{ and all } k \geqslant 1. \tag{2.23}$$

Second, provided that the weight distribution F has continuous density f, an application of the implicit function theorem gives us that the function $\epsilon_{kp}(x)$ is differentiable on $(0, kp\mu)$, and that its first derivative $\epsilon'_{kp}(x)$ is given by

$$\epsilon'_{kp}(x) = \frac{1}{kp\epsilon_{kp}(x)f(\epsilon_{kp}(x))} \qquad \text{if } 0 < x < kp\mu. \tag{2.24}$$

The proof of the regret bound then comes in two parts. In the next section we derive several estimates that have to do with the weight distribution belonging to the typical class and with $k \geqslant K$, while in Section 2.5.2 we estimate the gap $kprF(\epsilon_{kp}(x)) - \widehat{v}_k(x)$.

2.5.1 Preliminary Observations

When $k \geqslant K$ the properties that characterize typical weight distributions can be used to obtain general estimates that are crucial to our analysis. As a warm-up we obtain the following estimate on the mismatch between the probability of an item weight being smaller than the feasible threshold $\widehat{h}_k$ and the probability of the same weight being smaller than the consumption function ϵ_{kp}.

Lemma 2.3. *If the weight distribution F belongs to the typical class then there is $1 < M < \infty$ depending only on F such that*

$$\frac{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}{x} \leqslant M \quad \text{for all } x \in (0, c],\ p \in (0, 1], \text{ and all } k \geqslant K \equiv \left\lceil \frac{c}{p\int_0^{\bar{w}} wf(w)\,dw} \right\rceil. \tag{2.25}$$

In turn, we also have that

$$F(\epsilon_{kp}(x)) - F(\widehat{h}_k(x)) \leqslant \frac{M}{kp} \quad \text{for all } x \in [0, c],\ p \in (0, 1], \text{ and all } k \geqslant K. \tag{2.26}$$

Proof. The uniform bound (2.25) is essentially a restatement of inequality (2.21) in Lemma 2.2. If $x \in (0, c]$ and $k \geqslant K$, then we have that

$$\frac{x}{kp} \leqslant \frac{x}{Kp} \leqslant \frac{c}{Kp} \leqslant \int_0^{\bar{w}} w f(w)\, dw \leqslant \mu,$$

so the definition (2.2) of the consumption function $\epsilon_{kp}(x)$ and the equality (2.3) give us that

$$\epsilon_{kp}(x) \leqslant \bar{w} \quad \text{and} \quad \int_0^{\epsilon_{kp}(x)} w f(w)\, dw = \frac{x}{kp} \qquad \text{for all } k \geqslant K \text{ and all } x \in (0, c]. \tag{2.27}$$

The two observations in (2.27) together with the bound (2.21) in which we replace w with $\epsilon_{kp}(x)$ then imply that

$$\frac{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}{x} = \frac{\epsilon_{kp}(x)F(\epsilon_{kp}(x))}{\int_0^{\epsilon_{kp}(x)} u f(u)\, du} \leqslant M \quad \text{for all } k \geqslant K \text{ and } x \in (0, c],$$

concluding the proof of the uniform bound (2.25).

We now turn to inequality (2.26). If $x = 0$ then inequality (2.26) is obvious. Otherwise, if $x > 0$ we recall from (2.18) that $\widehat{h}_k(x) = \min\{x, \epsilon_{kp}(x)\}$, so the left-hand side of (2.26) is equal to 0 when $\epsilon_{kp}(x) \leqslant x < \infty$, and inequality (2.26) is again trivial. Instead, if $0 < x < \epsilon_{kp}(x)$, we obtain from (2.25) that

$$F(\epsilon_{kp}(x)) \leqslant \frac{Mx}{kp\epsilon_{kp}(x)} \leqslant \frac{M}{kp} \quad \text{for all } k \geqslant K \text{ and } 0 < x < \epsilon_{kp}(x),$$

concluding the proof of the lemma. □

In the same spirit of Lemma 2.3, we can also estimate the difference in the probability of selecting an upcoming item as a function of the number of remaining periods.

Lemma 2.4. *For $p \in (0, 1]$, all $x \in [0, c]$, and all $k \geqslant K$ we have that*

$$F(\epsilon_{(k+1)p}(x)) - F(\epsilon_{kp}(x)) \leqslant -\frac{x}{k(k+1)p\epsilon_{kp}(x)}.$$

Proof. For $k \geqslant K$ the equality (2.3) and the monotonicity (2.23) give us the representation

$$\frac{x}{kp} - \frac{x}{(k+1)p} = \int_{\epsilon_{(k+1)p}(x)}^{\epsilon_{kp}(x)} w f(w)\, dw, \qquad \text{for all } x \in [0, c].$$

If we now replace the integrand $wf(w)$ with the upper bound $\epsilon_{kp}(x)f(w)$ and rearrange, we obtain

$$\frac{x}{k(k+1)p} \leqslant \epsilon_{kp}(x)\left[F(\epsilon_{kp}(x)) - F(\epsilon_{(k+1)p}(x))\right],$$

completing the proof of the lemma. □

Typical weight distributions are also nice because one can tightly approximate the difference $F(\epsilon_{kp}(x))-F(\epsilon_{kp}(x-w))$ that accounts for the sensitivity in the remaining capacity of the probability of selecting the kth-to-last item. A formal estimate is given in the next proposition, and it constitutes a key step in our argument.

Proposition 2.2. *If $p \in (0,1]$ and if the weight distribution F belongs to the typical class, then there is a constant $1 < M < \infty$ depending only on F such that one has the inequality*

$$1 - \frac{F(\epsilon_{kp}(x-w))}{F(\epsilon_{kp}(x))} \leqslant \frac{w^2}{x^2}(1-M^{-1}) + \frac{w}{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))} \tag{2.28}$$

for all $w \in [0,x]$, $x \in (0,c]$, and all $k \geqslant K \equiv \left\lceil \frac{c}{p\int_0^w wf(w)\,dw} \right\rceil$.

The proof of Proposition 2.2 requires the following intermediate estimate.

Lemma 2.5 (Convexity upper bound)**.** *If $p \in (0,1]$ and if the weight distribution F has continuous density f then for all $k \geqslant K$, $x \in [0,c]$ and $y \in [0,1]$ we have the integral representation*

$$kpF(\epsilon_{kp}(x)) - kpF(\epsilon_{kp}(x(1-y))) = \int_{x(1-y)}^{x} \frac{1}{\epsilon_{kp}(u)}\,du. \tag{2.29}$$

Moreover, if the distribution F belongs to the typical class the map $x \mapsto \epsilon_{kp}(x)^{-1}$ is convex on $(0,c)$, so we also have the upper bound

$$kpF(\epsilon_{kp}(x)) - kpF(\epsilon_{kp}(x(1-y))) \leqslant \frac{xy}{2}\left[\frac{1}{\epsilon_{kp}(x)} + \frac{1}{\epsilon_{kp}(x(1-y))}\right]. \tag{2.30}$$

Proof. Since the weight distribution F has continuous density and $\frac{c}{p\mu} \leqslant K \leqslant k$, we recall from (2.24) the first derivative

$$\epsilon'_{kp}(x) = \frac{1}{kp\epsilon_{kp}(x)f(\epsilon_{kp}(x))} \qquad \text{for all } x \in (0,c).$$

The map $x \mapsto F(\epsilon_{kp}(x))$ is then differentiable on $(0, c)$, and one has that

$$(kpF(\epsilon_{kp}(x)))' = kp\epsilon'_{kp}(x)f(\epsilon_{kp}(x)) = \frac{1}{\epsilon_{kp}(x)} \qquad \text{for all } x \in (0, c).$$

In turn, the fundamental theorem of calculus tells us that for $y \in [0, 1]$ we have the integral representation

$$kpF(\epsilon_{kp}(x)) - kpF(\epsilon_{kp}(x(1-y))) = \int_{x(1-y)}^{x} \frac{1}{\epsilon_{kp}(u)}\, du,$$

proving the first assertion of the lemma.

To check the convexity of the map $x \mapsto \epsilon_{kp}(x)^{-1}$, we use the expression of the first derivative (2.24) one more time to obtain for $k \geqslant K$ that

$$\left(\frac{1}{\epsilon_{kp}(x)}\right)' = -\frac{\epsilon'_{kp}(x)}{\epsilon^2_{kp}(x)} = -\frac{1}{kp\epsilon^3_{kp}(x)f(\epsilon_{kp}(x))}.$$

If F belongs to the typical class and $k \geqslant K$ then the monotonicity condition (2.5) implies that the first derivative $(1/\epsilon_{kp}(x))'$ is non-decreasing on $(0, c)$, so the map $x \mapsto \epsilon_{kp}(x)^{-1}$ is convex. This convexity property then provides us with a linear majorant

$$m_{kp}(u) = \frac{u - x}{yx}\left(\frac{1}{\epsilon_{kp}(x)} - \frac{1}{\epsilon_{kp}((1-y)x)}\right) + \frac{1}{\epsilon_{kp}(x)}$$

such that

$$\frac{1}{\epsilon_{kp}(u)} \leqslant m_{kp}(u) \qquad \text{for all } u \in [(1-y)x, x].$$

The representation (2.29) and the integration of the majorant $m_{kp}(u)$ over $[(1-y)x, x]$ give us the upper bound (2.30), and the proof of the lemma follows. □

We now have all of the estimates we need to complete the proof of Proposition 2.2.

Proof. (Proof of Proposition 2.2.) If $w = 0$ then inequality (2.28) is trivial. Otherwise, for $K \leqslant k < \infty$ we consider the function $g_k : (0, c] \times (0, 1] \to \mathbb{R}$ given by

$$g_k(x, y) = \frac{1}{y^2}\left\{1 - \frac{F(\epsilon_{kp}\,(x(1-y)))}{F(\epsilon_{kp}(x))} - \frac{xy}{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}\right\},$$

and we note that inequality (2.28) follows by setting $y = w/x \leqslant 1$ and rearranging, provided that one has the uniform bound

$$g_k(x, y) \leqslant 1 - M^{-1} \qquad \text{for all } x \in (0, c],\ y \in (0, 1], \text{ and } k \geqslant K. \tag{2.31}$$

The function $g_k(x, y)$ is differentiable with respect to y for any given $x \in (0, c]$, and the y-derivative of $g_k(x, y)$ can be written as

$$\frac{\partial}{\partial y} g_k(x, y) = \frac{2}{y^3 kpF(\epsilon_{kp}(x))} \left\{ \frac{xy}{2} \left[\frac{1}{\epsilon_{kp}(x)} + \frac{1}{\epsilon_{kp}(x(1-y))} \right] - kpF(\epsilon_{kp}(x)) + kpF(\epsilon_{kp}(x(1-y))) \right\}.$$

Since $\frac{2}{y^3 kpF(\epsilon_{kp}(x))} \geqslant 0$, inequality (2.30) of Lemma 2.5 then tells us that the y-derivative of $g_k(x, y)$ is non-negative so that the map $y \mapsto g_k(x, y)$ is non-decreasing in y for any given $x \in (0, c]$. In turn, we have that

$$g_k(x, y) \leqslant g_k(x, 1) = 1 - \frac{x}{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))},$$

so inequality (2.31) follows from the uniform bound (2.25), and the proof of the proposition is now complete. □

2.5.2 Analysis of Residuals

To estimate the gap between the expected total reward collected by the reoptimized policy $\widehat{\pi} \in \Pi(n, c, p)$ and the prophet upper bound $nprF(\epsilon_{np}(c))$, we study appropriate residual functions. Specifically, we let

$$\rho_k(x) = kprF(\epsilon_{kp}(x)) - \widehat{v}_k(x) \qquad \text{for } x \in [0, c] \text{ and } 1 \leqslant k \leqslant n \tag{2.32}$$

be the *residual function* when there are k remaining periods and the level of remaining capacity is x. The residual function $\rho_k(x)$ is continuous and defined on a compact interval, so if we maximize with respect to x we obtain the *maximal residual*

$$\bar{\rho}_k = \max_{0 \leqslant x \leqslant c} \rho_k(x) \qquad \text{for } k \in [n]. \tag{2.33}$$

The second half of Theorem 2.1 is just a corollary of the following proposition, which verifies that the maximal residual $\bar{\rho}_n = O(\log n)$ as $n \to \infty$.

Proposition 2.3. *If the weight distribution F belongs to the typical class, then there is a constant $1 < M < \infty$ depending only on the distribution F, the arrival probability p, and the reward r such that the maximal residual*

$$\bar{\rho}_n = \max_{0 \leqslant x \leqslant c} \{nprF(\epsilon_{np}(x)) - \widehat{v}_n(x)\} \leqslant M + M \log n \qquad \text{for all } n \geqslant 1.$$

For the proof of this proposition we write the maximal residual $\bar{\rho}_n$ as a telescoping sum, and we obtain an appropriate upper bound for each summand. The upper bound follows from the following lemma.

Lemma 2.6. *If the weight distribution F belongs to the typical class, then there is a constant $1 < M < \infty$ that depends only on F and the reward r such that the difference*

$$\rho_{k+1}(x) - \bar{\rho}_k \leqslant \frac{M}{k+1} \qquad \text{for all } x \in [0, c] \text{ and all } k \geqslant K.$$

Proof. The residual function $\rho_k(x)$ defined in (2.32) provides us with an alternative representation for the value function $\widehat{v}_{k+1}(x)$ which gives us the expected total reward selected by policy $\widehat{\pi}$ with $k+1$ periods remaining and current knapsack capacity x. Specifically, if we substitute $\widehat{v}_k(x)$ with $kprF(\epsilon_{kp}(x)) - \rho_k(x)$ in the recursion (2.19), we then obtain that

$$\begin{aligned}\widehat{v}_{k+1}(x) =& \{1 - pF(\widehat{h}_{k+1}(x))\}\{kprF(\epsilon_{kp}(x)) - \rho_k(x)\} \\ &+ p \int_0^{\widehat{h}_{k+1}(x)} \{r + kprF(\epsilon_{kp}(x - w)) - \rho_k(x - w)\} f(w)\, dw.\end{aligned}$$

Next, if we replace the residuals $\rho_k(\cdot)$ with their maximal value $\bar{\rho}_k$ and rearrange, we obtain the lower bound

$$\begin{aligned} kprF(\epsilon_{kp}(x)) \quad &+ \quad prF(\widehat{h}_{k+1}(x)) + kp^2 r \int_0^{\widehat{h}_{k+1}(x)} \{F(\epsilon_{kp}(x - w)) - F(\epsilon_{kp}(x))\} f(w)\, dw \quad (2.34) \\ &\leqslant \quad \widehat{v}_{k+1}(x) + \bar{\rho}_k. \end{aligned}$$

In turn, the definition (2.32) of the residual function tells us that

$$\rho_{k+1}(x) - \bar{\rho}_k = (k+1)prF(\epsilon_{(k+1)p}(x)) - (\widehat{v}_{k+1}(x) + \bar{\rho}_k),$$

so if we replace the sum $\widehat{v}_{k+1}(x) + \bar{\rho}_k$ with its lower bound (2.34) and rearrange, we obtain the

upper bound

$$\begin{aligned}\rho_{k+1}(x)-\bar{\rho}_k &\leqslant pr\left\{(k+1)F(\epsilon_{(k+1)p}(x))-kF(\epsilon_{kp}(x))-F(\widehat{h}_{k+1}(x))\right\} \\ &\quad +kp^2rF(\epsilon_{kp}(x))\int_0^{\widehat{h}_{k+1}(x)}\left\{1-\frac{F(\epsilon_{kp}(x-w))}{F(\epsilon_{kp}(x))}\right\}f(w)\,dw.\end{aligned} \tag{2.35}$$

Next, we obtain from (2.28) that the integral that appears on the right-hand side of (2.35) satisfies the upper bound

$$\mathcal{I}_k(x)\equiv\int_0^{\widehat{h}_{k+1}(x)}\left\{1-\frac{F(\epsilon_{kp}(x-w))}{F(\epsilon_{kp}(x))}\right\}f(w)dw\leqslant\int_0^{\widehat{h}_{k+1}(x)}\left\{\frac{w^2}{x^2}(1-M^{-1})+\frac{w}{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}\right\}f(w)dw.$$

For $w\in[0,\widehat{h}_{k+1}(x)]$ we have the trivial bound $w^2\leqslant w\widehat{h}_{k+1}(x)$ so if we replace w^2 with its upper bound $w\widehat{h}_{k+1}(x)$ on the right-hand side above and integrate we obtain that there is $1<M<\infty$ depending only on F such that

$$\mathcal{I}_k(x)\leqslant\left[(1-M^{-1})\frac{\widehat{h}_{k+1}(x)}{x^2}+\frac{1}{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}\right]\int_0^{\widehat{h}_{k+1}(x)}wf(w)\,dw.$$

We now multiply both sides by $kp^2rF(\epsilon_{kp}(x))$ and simplify to obtain that

$$kp^2rF(\epsilon_{kp}(x))\mathcal{I}_k(x)\leqslant\left[kp^2rF(\epsilon_{kp}(x))(1-M^{-1})\frac{\widehat{h}_{k+1}(x)}{x^2}+\frac{pr}{\epsilon_{kp}(x)}\right]\int_0^{\widehat{h}_{k+1}(x)}wf(w)\,dw.$$

The definition of $\widehat{h}_{k+1}(x)=\min\{x,\epsilon_{(k+1)p}(x)\}$ and the monotonicity (2.23) tell us that $\widehat{h}_{k+1}(x)\leqslant\epsilon_{(k+1)p}(x)\leqslant\epsilon_{kp}(x)$, so we obtain a further upper bound if we replace the first $\widehat{h}_{k+1}(x)$ on the last right-hand side with $\epsilon_{kp}(x)$ and the second one with $\epsilon_{(k+1)p}(x)$. When we perform these replacements and recall the equality (2.3), we find that

$$kp^2rF(\epsilon_{kp}(x))\mathcal{I}_k(x)\leqslant\frac{r(1-M^{-1})}{k+1}\frac{kp\epsilon_{kp}(x)F(\epsilon_{kp}(x))}{x}+\frac{rx}{(k+1)\epsilon_{kp}(x)}.$$

If we now apply the uniform upper bound (2.25) to the first summand on the right-hand side, and rearrange, we obtain that

$$kp^2rF(\epsilon_{kp}(x))\mathcal{I}_k(x)\leqslant\frac{r(M-1)}{k+1}+\frac{rx}{(k+1)\epsilon_{kp}(x)}.$$

We now replace the last summand of (2.35) with the upper bound above and rearrange to obtain that

$$\rho_{k+1}(x) - \bar{\rho}_k \leqslant \frac{r(M-1)}{k+1} + kpr\left\{F(\epsilon_{(k+1)p}(x)) - F(\epsilon_{kp}(x)) + \frac{x}{k(k+1)p\epsilon_{kp}(x)}\right\}$$
$$+ pr\left\{F(\epsilon_{(k+1)p}(x)) - F(\widehat{h}_{k+1}(x))\right\}.$$

Here, Lemma 2.4 tells us that the second summand on the right-hand side is non-positive, and inequality (2.26) tells us that there is $1 < M < \infty$ depending only on F such that the difference $F(\epsilon_{(k+1)p}(x)) - F(\widehat{h}_{k+1}(x))$ is bounded above by $M/((k+1)p)$. When we assemble these observations, we finally find that

$$\rho_{k+1}(x) - \bar{\rho}_k \leqslant \frac{(2M-1)r}{k+1} \qquad \text{for all } x \in [0,c] \text{ and all } k \geqslant K,$$

concluding the proof of the lemma. □

We now have all of the tools we need to complete the proof of Proposition 2.3 that follows next.

Proof. (Proof of Proposition 2.3.) We write the maximal residual $\bar{\rho}_n$ in (2.33) as a telescoping sum and use the definition (2.32) of the residual function to obtain that

$$\bar{\rho}_n = \bar{\rho}_K + \sum_{k=K}^{n-1}\{\bar{\rho}_{k+1} - \bar{\rho}_k\} \leqslant K + \sum_{k=K}^{n-1}\{\bar{\rho}_{k+1} - \bar{\rho}_k\}.$$

Lemma 2.6 then tells us that

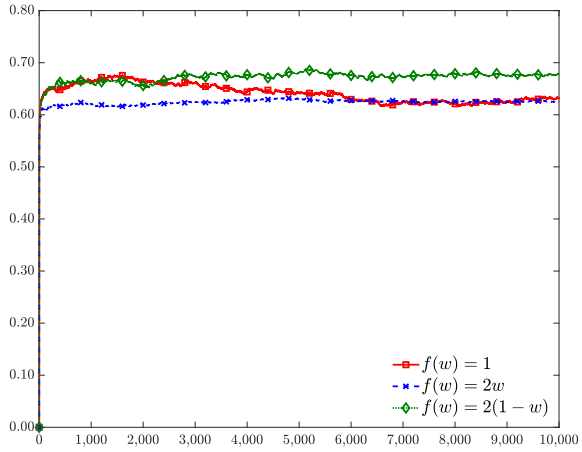
$$\bar{\rho}_{k+1} - \bar{\rho}_k \leqslant \frac{M}{k+1} \qquad \text{for all } K \leqslant k \leqslant n,$$

so when we combine the last two observations we obtain that there is a constant $1 < M < \infty$ that depends only on F, p, and r such that

$$\bar{\rho}_n \leqslant M + M\log n,$$

just as needed. □

Figure 2.1: **Gap between the prophet upper bound and offline sort for three weight distributions.**



Notes. Difference between the prophet upper bound, $nF(\epsilon_n(1))$, and the simulated average (with $100,000$ trials) of the offline solution, $R_n^*(1,1,1)$, for three different distributions on the unit interval: $f(w) = \mathbb{1}\{w \in (0,1)\}$, $f(w) = 2w\mathbb{1}\{w \in (0,1)\}$ and $f(w) = 2(1-w)\mathbb{1}\{w \in (0,1)\}$. In each case we take the arrival probability $p = 1$, the knapsack capacity $c = 1$, the reward $r = 1$, and we vary the number of periods $n \in \{1, 2, \ldots, 10000\}$. The chart suggests that the gap between the prophet upper bound and the simulated average of the offline solution does not grow with n.

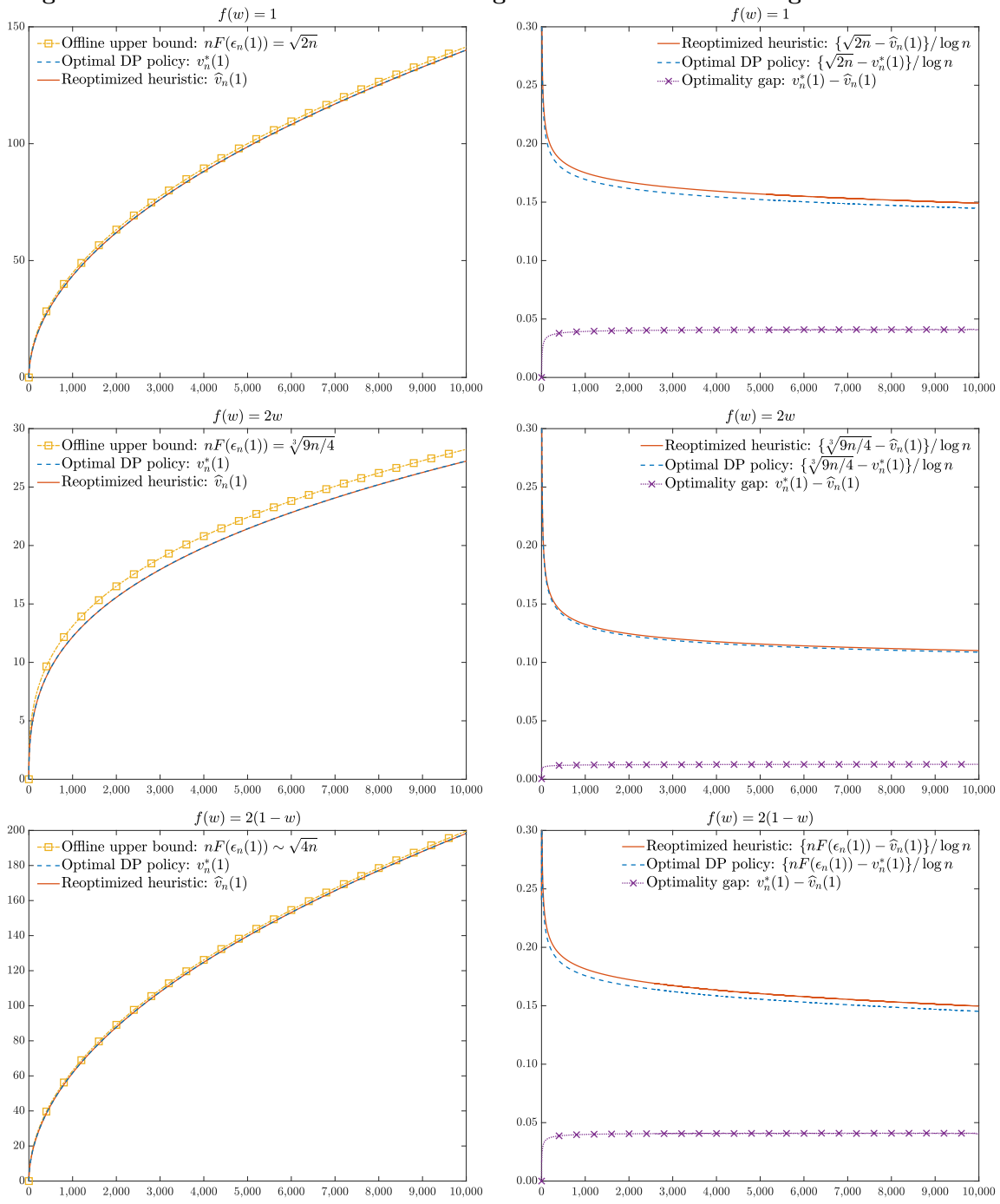
2.6 Numerical Experiments

Theorem 2.1 tells us that the regret of a dynamic and stochastic knapsack problem is at most logarithmic in n, provided that the weight distribution belongs to the typical class of Definition 2.1. While the actual order of the regret may—in principle—be smaller than what our bound predicts, we find numerically that this is not the case. In fact, we conjecture that the actual regret is $O(\log n)$ as $n \to \infty$ for most continuous weight distributions.

The work of Seksenbayev (2018) and Gnedin and Seksenbayev (2019) tells us that when the capacity, the reward, and the arrival probability are all equal to one, and the weight distribution is uniform on the unit interval, then the regret is asymptotic to $(\log n)/12$. In this section, we numerically investigate the actual order of the regret for two other weight distributions, while keeping the uniform as reference.

For our numerical examples, we solve the recursion (2.19) on a discretized state space with grid size 10^{-5} and obtain estimates for the reoptimized value function $\widehat{v}_n(\cdot)$ for $n \in \{1, \ldots, 10000\}$ and for different distributions F. On the same discretized state space and for the same weight distributions,

Figure 2.2: Value functions and scaled regret bounds for three weight distributions



Notes. The left plots display the prophet upper bound and the value functions of the optimal dynamic programming (DP) policy and of the reoptimized heuristic. The right plots show the regret bounds of the optimal policy and of the heuristic scaled by the logarithm of n, as well as the optimality gap. While the scaled regret bounds are bounded away from zero for large n, the optimality gap does not grow with n. Weights have densities on $(0, 1)$ respectively given by $f(w) = 1$, $f(w) = 2w$, and $f(w) = 2(1 - w)$. Capacity $c = 1$, arrival probability $p = 1$, and reward $r = 1$. Discretized state space with grid size 10^{-5}.

we also solve numerically the Bellman recursion

$$\begin{aligned} v_n^*(x) &= p(1-F(x))v_{n-1}^*(x) + p\int_0^x \max\{r + v_{n-1}^*(x-w), v_{n-1}^*(x)\} f(w)\,dw + (1-p)v_{n-1}^*(x) \\ &= (1-pF(x))v_{n-1}^*(x) + p\int_0^x \max\{r + v_{n-1}^*(x-w), v_{n-1}^*(x)\} f(w)\,dw \end{aligned} \tag{2.36}$$

with the initial condition $v_0^*(x) = 0$ for all $x \in [0, c]$, and we obtain estimates for the optimal value functions $v_n^*(\cdot)$ for $n \in \{1, \ldots, 10000\}$. Finally, we simulate the average of the offline solution $R_n^*(c, p, r)$ and compare all of our numerical estimates with the prophet upper bound $nprF(\epsilon_{np}(c))$. Based on our numerical experiments, we observe that:

(i) The gap $nprF(\epsilon_{np}(c)) - \mathbb{E}\left[R_n^*(c, p, r)\right]$ between the prophet upper bound and the offline solution is bounded by a constant that does not depend on n (see Figure 2.1);

(ii) The regret bound $nprF(\epsilon_{np}(c)) - \widehat{v}_n(c)$ for the reoptimized heuristic and the regret bound $nprF(\epsilon_{np}(c)) - v_n^*(c)$ for the optimal online policy grow logarithmically with n (Figure 2.2); and

(iii) The optimality gap $v_n^*(c) - \widehat{v}_n(c)$ is bounded by constant that is independent of n (Figure 2.2).

In turn, our numerical experiments suggests that the regrets (rather than the regret bounds) $\mathbb{E}\left[R_n^*(c, p, r)\right] - \widehat{v}_n(c)$ and $\mathbb{E}\left[R_n^*(c, p, r)\right] - v_n^*(c)$ respectively of the reoptimized heuristic and of the optimal online policy are also logarithmic in n. In our numerical work, we consider instances of the dynamic and stochastic knapsack problem with reward $r = 1$, arrival probability $p = 1$, and capacity $c = 1$. We vary item weights by considering the three densities supported on the unit interval given by $f(w) = 1$, $f(w) = 2w$ and $f(w) = 2(1-w)$ for $w \in (0, 1)$. The top left chart of Figure 2.2 plots the prophet upper bound $nF(\epsilon_n(1)) = \sqrt{2n}$ as well as the value function of the optimal policy, $v_n^*(1)$, and of the reoptimized heuristic, $\widehat{v}_n(1)$, when the weight distribution is uniform on $(0, 1)$. Instead, the top right chart depicts the respective regret bounds scaled by the logarithm of n, as well as the optimality gap. In the chart we see that the scaled regret bounds (top two lines) are bounded away from zero for large n, implying that the regret bounds grow logarithmically. In contrast, the optimality gap (bottom line) appears not to grow with n.

The plots in the middle row of Figure 2.2 point to the same set of observations when the weights have density $f(w) = 2w\mathbb{1}\{w \in (0, 1)\}$ and the prophet upper bound is $nF(\epsilon_n(1)) = \sqrt[3]{9n/4}$.

Finally, the bottom two charts of Figure 2.2 consider item weights that have density $f(w) = 2(1 - w)\mathbb{1}\{w \in (0, 1)\}$. In this case, the prophet upper bound cannot be expressed in closed form, but one can show that $nF(\epsilon_n(1)) \sim \sqrt{4n}$ as $n \to \infty$. Nevertheless, also for this weight distribution the numerical analysis suggests that the regrets of the optimal policy and of the heuristic are both logarithmic in n, and that the optimality gap can be bounded by a constant independent of n.

2.7 On Weight Distributions with Multiple Types

In this section, we discuss how our logarithm regret bound generalizes to dynamic and stochastic knapsack problems with equal rewards and with independent random weights that belong to one of $J < \infty$ different types. We consider a multinomial arrival process with parameters $\boldsymbol{p} \equiv (p_0, p_1, \ldots, p_J)$ where $p_j \in (0, 1]$ for all $j \in [J]$ and $p_0 = 1 - \sum_{i \in [J]} p_j \in [0, 1]$. Here, the parameter p_0 represents the probability of no item arriving (or, equivalently, the arrival probability of an item with arbitrarily large weight) and p_j, $j \in [J]$, is the arrival probability of an item with weight distribution F_j.

Upon arrival of an item the decision maker may see the type of the item or not. If the item types are *not* released, then she only sees the arriving weights that (conditional on an arrival occurring) are drawn from the mixture distribution

$$\widetilde{F}(w) = \frac{1}{1 - p_0} \sum_{j \in [J]} p_j F_j(w) \quad \text{for all } w \in [0, \infty).$$

If the weight distributions $F_1, F_2, \ldots, F_J$ are all typical (see Definition 2.1), then the mixture distribution $\widetilde{F}$ is also typical (see Section 2.4), and Theorem 2.1 immediately applies.

In contrast, if item types are revealed upon arrival, then the decision maker could use the type information to make better decisions. As we will see shortly, because the rewards are all equal, knowing the weight type of the arriving item makes no difference. The offline solution is still given by an algorithm that sorts items according to their realized weights (regardless of their types), and the optimal dynamic programming policy is a threshold policy that ignores weight types.

For the optimal offline solution, we can reinterpret this formulation so that items arrive according to a Bernoulli process with arrival probability $1 - p_0 = \sum_{j \in [J]} p_j$, have rewards equal to r and independent weights with distribution given by $\widetilde{F}$. The optimal offline solution $R_n^*(c, 1 - p_0, r)$ is

then given by the sorting algorithm (2.1), so if

$$\epsilon_{k(1-p_0)}(x) = \sup\left\{\epsilon \in [0,\infty) : \int_0^{\epsilon} w\, d\widetilde{F}(w) \leqslant \frac{x}{k(1-p_0)}\right\}, \tag{2.37}$$

then Proposition 2.1 gives us that

$$\mathbb{E}\left[R_n^*(c, 1-p_0, r)\right] \leqslant n(1-p_0) r \widetilde{F}(\epsilon_{n(1-p_0)}(c)), \tag{2.38}$$

and the prophet upper bound for weight distribution with multiple types follows.

To establish the independence on weight types of the optimal online solution when the rewards are all equal, we now examine the associated Bellman equation. We suppose that, with k periods to the end of the horizon, the remaining capacity is $x \in [0, c]$, the arriving item has weight type $j \in \{0, 1, \ldots, J\}$ (with $j = 0$ denoting a no arrival or, equivalently, an arrival with arbitrarily large weight), and we let $V_k(x, j)$ be the optimal expected reward to-go given the current state. The optimality principle of dynamic programming then tells us that the value function $V_k(x, j)$ satisfies the Bellman recursion

$$V_k(x,j) = (1-F_j(x))\sum_{\iota=0}^{J} p_\iota V_{k-1}(x,\iota) + \int_0^x \max\left\{r + \sum_{\iota=0}^{J} p_\iota V_{k-1}(x-w,\iota), \sum_{\iota=0}^{J} p_\iota V_{k-1}(x,\iota)\right\} dF_j(w), \tag{2.39}$$

with the initial condition $V_0(x, j) = 0$ for all $x \in [0, c]$ and all $j \in \{0, 1, \ldots, J\}$. Here, the first summand holds because with probability $1 - F_j(x)$ the arriving type-j item has weight that exceeds the current knapsack capacity and the decision maker must reject it. Thus, her expected reward to-go over the remaining $k-1$ periods is just given by the average over types of the value functions $V_{k-1}(x, \iota)$ for $\iota \in \{0, 1, \ldots, J\}$. Instead, with probability $F_j(x)$ the arriving type-j item can be selected and the decision maker chooses the action that yields the largest expected reward to-go. If the item has weight w then its selection yields $r + \sum_{\iota=0}^{J} p_\iota V_{k-1}(x-w, \iota)$, while its rejection gives $\sum_{\iota=0}^{J} p_\iota V_{k-1}(x, \iota)$. By integrating this against $F_j(\cdot)$ for $w \in [0, x]$, we obtain the second summand of (2.39). The value functions $V_k(x, j)$ are monotone increasing in x for each j and k, and one has that

$$H_k^*(x,j) = \sup\left\{w \in [0,x] : r + \sum_{\iota=0}^{J} p_\iota V_{k-1}(x-w,\iota) \geqslant \sum_{\iota=0}^{J} p_\iota V_{k-1}(x,\iota)\right\},$$

is the optimal threshold that identifies the largest type-j weight that can be selected when the current capacity is x and there are k periods remaining. Interestingly, one immediately has that

$H_k^*(x,j) = H_k^*(x,\iota)$ for all $j, \iota \in \{0,1,\ldots,J\}$ since all items have the same reward r and the expected rewards to-go of both actions are type independent. Because the optimal threshold policy ignores types, we can construct a heuristic that has the same property and use our earlier analysis to assess its performance. We recall the quantity $\epsilon_{k(1-p_0)}(x)$ in (2.37) and consider the type-independent threshold

$$\widehat{H}_k(x,j) = \min\{x, \epsilon_{k(1-p_0)}(x)\} \qquad \text{for all } j \in \{0,1,\ldots,J\} \text{ and } x \in [0,c].$$

If $\widehat{\pi}$ is the heuristic that uses the thresholds $\widehat{H}_n, \widehat{H}_{n-1}, \ldots, \widehat{H}_1$, and $R_n^{\widehat{\pi}}(c, 1-p_0, r)$ is the total reward that $\widehat{\pi}$ collects, then Proposition 2.3 tells us that there is a constant $1 < M < \infty$ depending only on $\widetilde{F}$, the arrival probability $1-p_0$, and the reward r such that

$$n(1-p_0)r\widetilde{F}(\epsilon_{n(1-p_0)}(c)) - M\log n \leqslant \mathbb{E}\left[R_n^{\widehat{\pi}}(c, 1-p_0, r)\right]. \tag{2.40}$$

If we combine the two bounds (2.38) and (2.40), we then have the corollary below.

Corollary 2.1 (Regret bound for weight distributions with multiple types)**.** *Consider a knapsack problem with capacity $0 \leqslant c < \infty$ and with items that arrive over $1 \leqslant n < \infty$ periods according to a multinomial process with parameters $\boldsymbol{p} \equiv (p_0, p_1, \ldots p_J)$ such that $1 - p_0 = \sum_{j\in[J]} p_j$, and where p_0 is the probability of no arrival. If the items have rewards all equal to r and type-dependent weights with continuous distributions $F_1, F_2, \ldots, F_J$ and mixture (conditional on an arrival occurring) given by*

$$\widetilde{F}(w) = \frac{1}{1-p_0}\sum_{j\in[J]} p_j F_j(w) \quad \text{for all } w \in [0,\infty),$$

then

$$\mathbb{E}\left[R_n^*(c, 1-p_0, r)\right] \leqslant n(1-p_0)r\widetilde{F}(\epsilon_{n(1-p_0)}(c)).$$

Furthermore, there is a feasible online policy $\widehat{\pi}$ such that if the weights are independent and their distributions $F_1, \ldots, F_J$ belong to the typical class then there is a constant M depending only on $\widetilde{F}$, p_0, and r for which

$$n(1-p_0)r\widetilde{F}(\epsilon_{n(1-p_0)}(c)) - M\log n \leqslant \mathbb{E}\left[R_n^{\widehat{\pi}}(c, 1-p_0, r)\right].$$

In turn, if the weights are independent and $F_1, \ldots, F_J$ all belong to the typical class, then we have the regret bound

$$\mathbb{E}\left[R_n^*(c, 1-p_0, r)\right] - \mathbb{E}\left[R_n^{\widehat{\pi}}(c, 1-p_0, r)\right] \leqslant M\log n.$$

We note here that the key assumption that makes our analysis carry over to weight distributions with multiple types is that the rewards are all equal. If one were to allow for type-dependent rewards, then the optimal offline solution would *not* be given by the offline-sort algorithm (2.1) and the optimal online solution would *not* be given by type-independent thresholds. While one would still have a Bellman recursion analogous to (2.39), it is unclear how type-dependent rewards would affect our regret estimates, and we leave this interesting open problem for future research.

2.8 Concluding Remarks

In this chapter we study the dynamic and stochastic knapsack problem with equal rewards and independent random weights with common continuous distribution F. We prove that—under some mild regularity conditions on the weight distribution—the regret is, at most, logarithmic in n. In particular, we show that this regret bound is attained by a reoptimized heuristic that can be expressed explicitly and that provides a key analytical connection with the offline solution.

Two questions stem naturally from our analysis. The first one entails the difference in performance between the reoptimized heuristic and the optimal online policy. Based on our numerical experiments, we conjecture that

$$\max_{\pi\in\Pi(n,c,p)} \mathbb{E}\left[R_n^{\pi}(c,r,p)\right] = \mathbb{E}\left[R_n^{\hat{\pi}}(c,r,p)\right] + O(1) \tag{2.41}$$

for all $n \geqslant 1$ and for a large class of weight distributions. However, it is well-known that the optimal policy often lacks of desirable structural properties, so proving (2.41) is unlikely to be easy. The second question has to do with the performance of the offline-sort algorithm. Here, the numerical evidence suggests that

$$\mathbb{E}\left[R_n^{*}(c,r,p)\right] = nprF(\epsilon_{np}(c)) + O(1)$$

for all $n \geqslant 1$ and most continuous weight distributions F.

Resolving the two conjectures above would imply that the regret cannot be $o(\log n)$ as $n \to \infty$ for most continuous weight distributions, and that $O(\log n)$ as $n \to \infty$ correctly quantifies the informational advantage that the prophet has over the sequential decision maker. This is in contrast with some other dynamic and stochastic knapsack problems in which the sequential decision maker does essentially as well as the prophet (see Section 2.1). It also suggests that when items have random weights, then the design of near-optimal heuristics requires more care than usual.

Chapter 3

An Adaptive $O(\log n)$-Optimal Policy for the Sequential Selection of a Monotone Subsequence from a Random Sample

In the problem of *sequential* (online) selection of a *monotone increasing* subsequence, a decision maker observes sequentially a sequence of independent non-negative random variables $X_1, X_2, \ldots$ with common continuous distribution F and seeks to construct a monotone subsequence

$$X_{\tau_1} \leqslant X_{\tau_2} \leqslant \cdots \leqslant X_{\tau_j} \tag{3.1}$$

where the indices $1 \leqslant \tau_1 < \tau_2 < \cdots < \tau_j$ are stopping times with respect to the σ-fields $\mathcal{F}_i = \sigma\{X_1, X_2, \ldots, X_i\}$, $1 \leqslant i < \infty$, and the trivial σ-field $\mathcal{F}_0$. Since the indices are required to be possible values of stopping times, all selection/rejection decisions are terminal. That is, if the decision maker chooses not to select the value X_i at time i, then that value is lost forever. Similarly, if X_i is selected at time i, then that selection cannot be changed in the future. In general, the stopping times can be chosen to optimize different objective functions, and two main approaches have been considered in the literature. In the first, the decision maker seeks to maximize the expected number of selected elements when n are sequentially revealed (Samuels and Steele, 1981). In contrast, in the second approach the decision maker's objective is to minimize the expected time it takes to construct a monotone subsequence with n elements (Arlotto et al., 2016). Here, we confine our attention to the first — more classical — approach.

We then call a sequence of stopping times $1 \leqslant \tau_1 < \tau_2 < \cdots < \tau_j \leqslant n$ such that (3.1) holds a *feasible policy*, and we denote the set of all such policies by $\Pi(n)$. For any $\pi \in \Pi(n)$, we then let $L_n(\pi)$ be the random variable that counts the number of selections made by policy π for the sample $\{X_1, X_2, \ldots, X_n\}$. That is,

$$L_n(\pi) = \max\{j : X_{\tau_1} \leqslant X_{\tau_2} \leqslant \cdots \leqslant X_{\tau_j} \text{ where } 1 \leqslant \tau_1 < \tau_2 < \cdots < \tau_j \leqslant n\}.$$

Samuels and Steele (1981) first studied this selection problem and found that for each $n \geqslant 1$ there

This chapter is written under the supervision of Prof. Alessandro Arlotto and Prof. Yehua Wei. The results presented here are also in a joint paper Arlotto, Wei, and Xie (2018), published in *Random Structures & Algorithms*.

is a unique policy $\pi^* \in \Pi(n)$ such that

$$\mathbb{E}\left[L_n(\pi^*)\right] = \sup_{\pi \in \Pi(n)} \mathbb{E}\left[L_n(\pi)\right], \tag{3.2}$$

and for such optimal policies one has the asymptotic estimate

$$\mathbb{E}\left[L_n(\pi^*)\right] \sim (2n)^{1/2} \quad \text{as } n \to \infty. \tag{3.3}$$

Over the last few decades the understanding of policy π^* has substantially evolved. For instance, by formulating (3.2) as a finite-horizon Markov decision problem one sees that the optimal policy π^* is characterized by time and state dependent acceptance intervals. Furthermore, the asymptotic estimate (3.3) was refined with the much tighter bounds

$$(2n)^{1/2} - O(\log n) \leqslant \mathbb{E}\left[L_n(\pi^*)\right] \leqslant (2n)^{1/2} \qquad \text{as } n \to \infty. \tag{3.4}$$

The upper bound in (3.4) was first discovered by Bruss and Robertson (1991) while studying the maximal number of elements of a random sample whose sum is less than a specified value. The analysis was instigated by the work of Coffman et al. (1987), and it now represents one of the early steps into the domain of resource-dependent branching processes (see, e.g., Bruss and Duerinckx, 2015). The result of Bruss and Robertson (1991) is actually quite rich; recent extensions and applications are discussed in Steele (2016). The upper bound in (3.4) also appeared in Gnedin (1999) who considered the sequential selection of a monotone increasing subsequence from a random sample with random size.

The $O(\log n)$ lower bound in (3.4) is much more recent. It first appeared in the work of Bruss and Delbaen (2001) who studied the mean-optimal sequential selection of a monotone increasing subsequence when the observations $X_1, X_2, \ldots$ are revealed at the arrival epochs of a unit-rate Poisson process on $[0, n]$. While the Bruss and Delbaen (2001) result provides compelling evidence that a similar bound should also hold for the discrete-time formulation we consider here — i.e. the formulation in which the observations are revealed at the times $1, 2, \ldots, n$ — the sequential nature of the two selection processes makes the result of Bruss and Delbaen (2001) not immediately applicable. The connection between the continuous-time formulation of Bruss and Delbaen (2001) and the discrete-time optimization (3.2) was then argued by Arlotto et al. (2015) who used the concavity of the map $n \mapsto \mathbb{E}\left[L_n(\pi^*)\right]$ and the $O(\log n)$-bound of Bruss and Delbaen (2001) to ultimately confirm the lower bound in (3.4).

After a careful analysis, Bruss and Delbaen (2004) proved that the mean-optimal number of monotone increasing selections with Poisson-many observations is asymptotically normal after centering around $(2n)^{1/2}$ and scaling by $3^{-1/2}(2n)^{1/4}$. Arlotto et al. (2015) showed that the same asymptotic limit also holds for the discrete-time problem with n observations so, in summary, we now know that

$$\frac{3^{1/2}\{L_n(\pi^*) - (2n)^{1/2}\}}{(2n)^{1/4}} \Longrightarrow N(0,1), \qquad \text{as } n \to \infty. \tag{3.5}$$

However, the analyses of Bruss and Delbaen (2001, 2004) and Arlotto et al. (2015) do not address whether there is a simple *adaptive* online policy — i.e., a policy that depends on the value of the last selection and on the number of observations that are yet to be seen — that is $O(\log n)$ optimal. The works of Rhee and Talagrand (1991) and Arlotto and Steele (2011) tell us that the best non-adaptive policy is $O(n^{1/4})$ optimal, but this optimality gap is too crude. For instance, the expected number of monotone increasing selections made by the best non-adaptive policy cannot even be used to center the random variable $L_n(\pi^*)$ around $(2n)^{1/2}$ in the weak law (3.5).

In this chapter, we construct a *simple adaptive online policy* $\widehat{\pi}$ that is $O(\log n)$ optimal. The policy is characterized by a sequence of functions $\widehat{\eta}_n, \widehat{\eta}_{n-1}, \ldots, \widehat{\eta}_1$ such that if the value of the last selection up to and including time i is s and if $k = n - i$ observations remain to be seen then the value X_{i+1} is selected if and only if X_{i+1} falls in the *acceptance interval* $[s, \widehat{\eta}_{n-i}(s)]$. In terms of the stopping times, the policy $\widehat{\pi}$ corresponds to setting $\widehat{\tau}_0 = 0$, $X_{\widehat{\tau}_0} = 0$, and then defining the stopping times $\widehat{\tau}_1 < \widehat{\tau}_2 < \cdots < \widehat{\tau}_j$ recursively as

$$\widehat{\tau}_j = \min\{\widehat{\tau}_{j-1} < i \leqslant n : X_i \in [X_{\widehat{\tau}_{j-1}}, \widehat{\eta}_{n-i+1}(X_{\widehat{\tau}_{j-1}})]\} \quad \text{for } 1 \leqslant j \leqslant n,$$

with the convention that if the set of indices on the right-hand side is empty, then $\widehat{\tau}_j = \infty$. The random variable $L_n(\widehat{\pi})$ then denotes the number of monotone increasing selections made by policy $\widehat{\pi}$, and the expected value of $L_n(\widehat{\pi})$ satisfies the two bounds given in the next theorem.

Theorem 3.1 ($O(\log n)$-Optimal Policy)**.** *For each $n \geqslant 1$, there is a simple adaptive online policy $\widehat{\pi}$ such that*

$$(2n)^{1/2} - 2\{\log(n) + 1\} \leqslant \mathbb{E}\left[L_n(\widehat{\pi})\right] \leqslant \mathbb{E}\left[L_n(\pi^*)\right] \leqslant (2n)^{1/2}. \tag{3.6}$$

We discuss the structure of policy $\widehat{\pi}$ and of the functions $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$ in Section 3.1, and then we turn to the proof of Theorem 3.1. The upper bound in (3.6) immediately follows from

(3.4) and the optimality of policy π^*, but our analysis requires a generalization of the upper bound (3.4) which we study in Section 3.2. Then in Section 3.3, we establish the equivalence between the sequential monotone subsequence selection problem and a special instance of the dynamic and stochastic knapsack problem with equal rewards of Chapter 2. With this equivalence, we can directly apply our previous results to prove the lower bound (3.6). Finally, in Section 3.4 we make concluding remarks and underscore some open problems.

3.1 Policy $\widehat{\pi}$ and Its Value Function

For any feasible policy π and any continuous distribution F, we see that the number of selections made by π for the sample $\{X_1, X_2, \ldots, X_n\}$ is unchanged if we replace each X_i by its monotone transformation $F^{-1}(X_i)$. Thus, we can assume without loss of generality that the X_i's are uniformly distributed on $[0, 1]$. Next, we let

$$\widehat{\eta}_k(s) = \min\left\{s + [2k^{-1}(1-s)]^{1/2},\, 1\right\} \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1, \tag{3.7}$$

and we use the sequence of functions $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$ to construct appropriate acceptance intervals. Specifically, if s denotes the value of the last selection when k observations are yet to be seen and x is the kth-to-last presented value, then x is selected as element of the subsequence that is under construction if and only if $x \in [s, \widehat{\eta}_k(s)]$.

We now define the *critical value* $s_k = \max\{1-2k^{-1}, 0\}$ and we note that for $s \in [0, s_k]$ the decision maker is *conservative* and selects the kth-to-last value x if and only if it is within $\{2k^{-1}(1-s)\}^{1/2}$ of the most recent selection s. On the other hand, if $s \in [s_k, 1]$ the decision maker is *greedy* and accepts any kth-to-last value x that is larger than the most recent selection s.

If k denotes the number of observations that are yet to be seen and s is the value of the most recent selection, then we let $\widehat{\nu}_k(s)$ denote the expected number of monotone increasing selections made by the acceptance interval policy characterized by the functions $\widehat{\eta}_k, \widehat{\eta}_{k-1}, \ldots, \widehat{\eta}_1$. The functions $\{\widehat{\nu}_k : 1 \leqslant k \leqslant n\}$ are the value functions associated with policy $\widehat{\pi}$ and they can be obtained recursively. Specifically, if $\widehat{\nu}_0(s) = 0$ for all $s \in [0, 1]$, then for $k \geqslant 1$ we have the recursion

$$\widehat{\nu}_k(s) = \{1 - \widehat{\eta}_k(s) + s\}\widehat{\nu}_{k-1}(s) + \int_s^{\widehat{\eta}_k(s)} \{1 + \widehat{\nu}_{k-1}(x)\}\, dx. \tag{3.8}$$

To see why this recursion holds, we condition on the kth-to-last uniform random value. With

probability $1-\widehat{\eta}_k(s)+s$ the newly presented value x does not fall in the acceptance interval $[s, \widehat{\eta}_k(s)]$. In this case no selection is made and we are left with $k-1$ remaining observations and with the value of the most recent selection s unchanged. This amounts to an expected number of remaining selections equal to $\widehat{\nu}_{k-1}(s)$ and it justifies the first summand of our recursion (3.8). On the other hand, with probability $\widehat{\eta}_k(s) - s$ the newly presented value x falls in the acceptance interval, and we obtain a reward of one for selecting x plus the expected number of remaining selections over the next $k-1$ observations when the value of the most recent selection changes to x. Integrating this over all $x \in [s, \widehat{\eta}_k(s)]$ gives us the second summand of the recursive equation (3.8). The value functions $\{\widehat{\nu}_k : 1 \leqslant k < \infty\}$ are all continuous on $[0,1]$ and their behavior is well summarized by the following theorem.

Theorem 3.2 ($\widehat{\nu}_k$ Bounds)**.** *For all $k \geqslant 1$ and all $s \in [0,1]$ one has that*

$$\{2k(1-s)\}^{1/2} - 2\{\log(k)+1\} \leqslant \widehat{\nu}_k(s) \leqslant \{2k(1-s)\}^{1/2}. \tag{3.9}$$

Since policy $\widehat{\pi}$ is characterized by the thresholds $\widehat{\eta}_n, \widehat{\eta}_{n-1}, \ldots, \widehat{\eta}_1$ and by the initial state $s=0$, one then has the equivalence

$$\mathbb{E}\left[L_n(\widehat{\pi})\right] = \widehat{\nu}_n(0) \qquad \text{for all } n \geqslant 1.$$

Hence, Theorem 3.1 is an immediate corollary of Theorem 3.2, but the upper bound (3.9) is a refinement of (3.4) to an arbitrary initial state. In fact the same upper bound holds for all feasible policies based on acceptance intervals, including the optimal one.

The estimates in Theorem 3.2 also allow us to provide some intuition for our choice of the threshold functions $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$ in (3.7). For every $k \geqslant 1$ and $s \in [0,1]$ the threshold functions aim to balance the expected reward to-go $\widehat{\nu}_{k-1}(s)$ that one obtains when skipping the kth-to-last observation x, and the reward $1 + \widehat{\nu}_{k-1}(x)$ that one earns when selecting the kth-to-last value x. Since $\widehat{\nu}_{k-1}(s) \approx \{2(k-1)(1-s)\}^{1/2}$, one can solve the equation

$$\{2(k-1)(1-s)\}^{1/2} = 1 + \{2(k-1)(1-\widehat{x})\}^{1/2} \tag{3.10}$$

to find the largest value of x that makes selecting the current value worthwhile. Equation (3.10) then tells us that

$$\widehat{x} = s + [2(k-1)^{-1}(1-s)]^{1/2} - [2(k-1)]^{-1} \approx s + [2k^{-1}(1-s)]^{1/2} - [2k]^{-1},$$

so our choice (3.7) accounts for the first two terms of the approximation of the solution of equation (3.10). The truncation in (3.7) then ensures that all the thresholds $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$ are feasible. At the end of the next section we use an optimization argument to provide further intuition for our choice of the threshold functions.

3.2 A Refined Prophet Upper Bound

In this section, we prove the upper bound (3.9) by showing that it holds for all policies that are based on acceptance intervals. The adaptive policy $\widehat{\pi}$ and the unique optimal policy π^* both have this property. The argument we provide draws substantially from the earlier analyses of Gnedin (1999) and Bruss and Delbaen (2001), but it takes advantage of the flexibility that comes from allowing for an arbitrary initial state. Specifically, here we assume that the first subsequence element can be selected only if it is larger than an arbitrary value $s \in [0, 1]$. In contrast, in the classical formulation one always takes the initial state $s = 0$.

For any $k \geqslant 1$, an arbitrary acceptance interval policy π_k is given by a sequence $\eta_k, \eta_{k-1}, \ldots, \eta_1$ of functions such that

$$s \leqslant \eta_j(s) \leqslant 1 \qquad \text{for all } 1 \leqslant j \leqslant k \text{ and all } s \in [0, 1].$$

If s denotes the initial state or the value of the last observation selected prior to being presented with the j-to-last value x, then x is selected if and only if $x \in [s, \eta_j(s)]$. Next, for any $s \in [0, 1]$ we set $M_0 = s$ and we let M_i denote the maximum between M_0 and the largest of the elements of the subsequence that have been selected up to and including time i. The number of selections from $\{X_1, X_2, \ldots, X_k\}$ when $M_0 = s$ is then given by the random variable

$$L_k(\pi_k, s) = \sum_{i=1}^{k} \mathbb{1}\left\{X_i \in [M_{i-1}, \eta_{k-i+1}(M_{i-1})]\right\}.$$

If we now take expectations on both sides and use the Cauchy-Schwarz inequality to estimate the sum of the products $1 \cdot \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right]$ for $1 \leqslant i \leqslant k$, we obtain

$$\begin{aligned} \mathbb{E}\left[L_k(\pi_k, s)\right] &= \sum_{i=1}^{k} 1 \cdot \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right] \qquad (3.11) \\ &\leqslant k^{1/2} \left\{ \sum_{i=1}^{k} \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right]^2 \right\}^{1/2}. \end{aligned}$$

The definition of M_i as the maximum between M_0 and the largest of the elements of the subsequence that have been selected up to and including time i tells us that

$$M_i = \begin{cases} M_{i-1} & \text{if } X_i \notin [M_{i-1}, \eta_{k-i+1}(M_{i-1})] \\ X_i & \text{if } X_i \in [M_{i-1}, \eta_{k-i+1}(M_{i-1})] \end{cases}$$

and, because X_i is uniformly distributed on the unit interval, we have the identity

$$\mathbb{E}\left[M_i - M_{i-1} \,|\, \mathcal{F}_{i-1}\right] = \int_{M_{i-1}}^{\eta_{k-i+1}(M_{i-1})} (x - M_{i-1})\, dx = \frac{1}{2}\left(\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right)^2.$$

Taking the total expectation then gives

$$\mathbb{E}\left[\left(\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right)^2\right] = 2\{\mathbb{E}\left[M_i\right] - \mathbb{E}\left[M_{i-1}\right]\},$$

so a second application of the Cauchy-Schwarz inequality implies the upper bound

$$\mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right]^2 \leqslant 2\{\mathbb{E}\left[M_i\right] - \mathbb{E}\left[M_{i-1}\right]\}.$$

If we now sum over $1 \leqslant i \leqslant k$ and recall that $\mathbb{E}\left[M_k\right] \leqslant 1$ and $M_0 = s$, we obtain from telescoping that

$$\sum_{i=1}^{k} \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right]^2 \leqslant 2\{\mathbb{E}\left[M_k\right] - \mathbb{E}\left[M_0\right]\} \leqslant 2(1-s). \tag{3.12}$$

This last inequality and the bound in (3.11) finally give us that

$$\mathbb{E}\left[L_k(\pi_k, s)\right] \leqslant \{2k(1-s)\}^{1/2} \qquad \text{for all } k \geqslant 1 \text{ and all } s \in [0,1]$$

and complete the proof of the upper bound (3.9).

The same upper bound can also be obtained by formulating a simple optimization problem that provides further insight into our choice of the adaptive thresholds $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$. We consider the optimization problem

$$w^* = \max_{d_1,\ldots,d_k} \sum_{i=1}^{k} d_i \tag{3.13}$$

$$\text{s.t. } \sum_{i=1}^{k} d_i^2 \leqslant 2(1-s),$$

and we obtain from inequality (3.12) that $d_i = \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right]$ is a feasible solution. This feasible solution has objective

$$\sum_{i=1}^{k} \mathbb{E}\left[\eta_{k-i+1}(M_{i-1}) - M_{i-1}\right] = \mathbb{E}\left[L_k(\pi_k, s)\right],$$

so we have that the optimal objective value of (3.13) is an upper bound for $\mathbb{E}\left[L_k(\pi_k, s)\right]$, and that (3.13) is a relaxation of the sequential monotone subsequence problem (3.2).

The optimal value of (3.13) can be easily estimated. The problem has a linear objective function and convex constraints; by the Karush–Kuhn–Tucker (KKT) conditions, its optimal solution $(d_1^*, d_2^*, \ldots, d_k^*)$ is given by

$$d_i^* = \left[2k^{-1}(1-s)\right]^{1/2} \qquad \text{for all } 1 \leqslant i \leqslant k,$$

and its optimal value is

$$w^* = \left\{2k(1-s)\right\}^{1/2} \qquad \text{for all } k \geqslant 1 \text{ and } s \in [0,1].$$

It then follows that the adaptive thresholds $\{\widehat{\eta}_k : 1 \leqslant k < \infty\}$ defined by (3.7) are *reoptimized* thresholds that use the optimal solution of the relaxation (3.13) for any given $k \geqslant 1$ and $s \in [0,1]$. Of course, a small nuisance arises, and one needs to make sure that the reoptimized thresholds are indeed feasible. Since there are values $s \in [0,1]$ such that $s + \left[2k^{-1}(1-s)\right]^{1/2} > 1$, one needs the truncation introduced in (3.7) to obtain thresholds feasible. In the next section, we connect the sequential monotone subsequence selection problem to the dynamic and stochastic knapsack problem of Chapter 2, and apply the established results to prove the lower bound in Theorem 3.2.

3.3 Equivalence with the Dynamic and Stochastic Knapsack Problem with Equal Rewards

As we mentioned in Chapter 2, the problem of the sequential selection of a monotone subsequence from a random sample is equivalent to a special instance of the dynamic and stochastic knapsack problem with equal rewards. More specifically, the equivalent special instance is the one in which the items arrive deterministically in every period, have unitary rewards and independent random weights drawn from the uniform distribution on $[0,1]$, and the knapsack has unit capacity. Such equivalence can be easily seen with one change of variable once one writes down the Bellman

equations for the two problems. Specifically, for the problem of sequential increasing subsequence selection, the Bellman equation is given by the recursion

$$\nu_k^*(s) = s\nu_{k-1}^*(s) + \int_s^1 \max\{1 + \nu_{k-1}^*(x), \nu_{k-1}^*(s)\}\, dx, \quad \text{for } k \geqslant 1, \tag{3.14}$$

and by the boundary condition $\nu_0^*(s) \equiv 0$ for all $s \in [0,1]$. The value function $\nu_k^*(s)$ represents the expected number of monotone increasing selections made by the optimal policy when there are k values yet to be seen and the last selected subsequence element is s. The optimality of the Bellman equation (3.14) can be verified by conditioning on the kth-to-last uniform random value x. On the one hand, with probability s this newly realized value is smaller than the last selected value s. In this case, no selection can be made and one is left with $k-1$ random values yet to observe and the last selected value s is unchanged. This yields the first term on the right-hand side of (3.14). On the other hand, if the uniform random value $x \geqslant s$, then one chooses to select or to reject this value. A selection of x yields one element in the selected subsequence and an expected number of future selections $\nu_{k-1}^*(x)$ since there are $k-1$ random values and the last selected value now is x. On the contrary, a rejection yields no selected element and an expected number of future selections $\nu_{k-1}^*(s)$ since the last selected value s is unchanged. By integrating the maximum of the two yields over the interval $[s,1]$, we recover the second term on the right-hand side of the optimality equation (3.14).

If we now consider a dynamic and stochastic knapsack problem with initial capacity $c = 1$, items that arrive in each period with probability $p = 1$, rewards $r = 1$, and random weights that are drawn uniformly on the unit interval, then we obtain from (2.36) that the Bellman equation is given by the recursion

$$v_k^*(s) = (1-s)v_{k-1}^*(s) + \int_0^s \max\{1 + v_{k-1}^*(s-w), v_{k-1}^*(s)\}\, dw, \quad \text{for } k \geqslant 1, \tag{3.15}$$

and by the boundary condition $v_0^*(s) \equiv 0$ for all $s \in [0,1]$.

The equivalence between the above two problems is summarized in the following proposition.

Proposition 3.1. *For all $k \geqslant 1$ and all $s \in [0,1]$, one has that*

$$\nu_k^*(s) = v_k^*(1-s).$$

Proof. The two recursions have the same boundary condition $\nu_0^*(s) = v_0^*(1-s) = 0$ for all $s \in [0,1]$,

so we have the induction case. Now, we take as induction hypothesis that

$$\nu^*_{k-1}(s) = v^*_{k-1}(1-s) \quad \text{for all } s \in [0,1],$$

and we prove that $\nu^*_k(s) = v^*_k(1-s)$ for all $s \in [0,1]$. By replacing s with $1-s$ in (3.15) we obtain that

$$v^*_k(1-s) = (1-s)v^*_{k-1}(1-s) + \int_0^{1-s} \max\{1 + v^*_{k-1}(1-s-w), v^*_{k-1}(1-s)\}\, dw.$$

On the right-hand side of this last equation, we can now apply the induction hypothesis to replace $v^*_{k-1}(1-s)$ with $\nu^*_{k-1}(s)$ and to replace $v^*_{k-1}(1-s-w)$ with $\nu^*_{k-1}(s+w)$. We then obtain that

$$v^*_k(1-s) = (1-s)\nu^*_{k-1}(s) + \int_0^{1-s} \max\{1 + \nu^*_{k-1}(s+w), \nu^*_{k-1}(s)\}\, dw.$$

If we now change variable and let $w = x - s$ to this last right-hand side, we obtain that

$$v^*_k(1-s) = (1-s)\nu^*_{k-1}(s) + \int_s^1 \max\{1 + \nu^*_{k-1}(x), \nu^*_{k-1}(s)\}\, dx.$$

By recalling the Bellman equation (3.14) for the sequential monotone subsequence selection problem, we finally obtain that $\nu^*_k(s) = v^*_k(1-s)$ for all $s \in [0,1]$, just as needed. □

In addition to the equivalence between the optimal policies of the two problems, one also has the equivalence between the adaptive policy $\widehat{\pi}$ of Section 2.3 and the adaptive policy $\widehat{\pi}$ of Section 3.1. To see this equivalence, let us recall that for the special instance of the dynamic and stochastic knapsack problem, we know from (2.2) and (2.18) that the threshold of policy $\widehat{\pi}$ for the knapsack problem is given by

$$\widehat{h}_k(s) = \min\{s, (2k^{-1}s)^{1/2}\} \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1. \tag{3.16}$$

On the other hand, the threshold of policy $\widehat{\pi}$ for the sequential monotone subsequence selection problem in (3.7) is given by

$$\widehat{\eta}_k(s) = \min\left\{s + [2k^{-1}(1-s)]^{1/2},\, 1\right\} \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1, \tag{3.17}$$

and one can easily see that the two thresholds (3.16) and (3.17) satisfy the following relation

$$\widehat{h}_k(1-s) = \widehat{\eta}_k(s) - s \tag{3.18}$$

The identity above provides the essential link to connect the value function of the policy $\widehat{\pi}$ for the dynamic and stochastic knapsack problem in (2.19) and given by

$$\widehat{v}_k(s) = (1 - \widehat{h}_k(s))\widehat{v}_{k-1}(s) + \int_0^{\widehat{h}_k(s)} (1 + \widehat{v}_{k-1}(s-w))\, dw, \tag{3.19}$$

with the value function of policy $\widehat{\pi}$ for the sequential monotone subsequence selection problem, which we recall from (3.8), as

$$\widehat{\nu}_k(s) = \{1 - \widehat{\eta}_k(s) + s\}\widehat{\nu}_{k-1}(s) + \int_s^{\widehat{\eta}_k(s)} \{1 + \widehat{\nu}_{k-1}(x)\}\, dx. \tag{3.20}$$

Just as in Proposition 3.1, we see that the two recursions (3.19) and (3.20) are equivalent in the following sense.

Proposition 3.2. *For all $k \geqslant 1$ and all $s \in [0,1]$, one has that*

$$\widehat{\nu}_k(s) = \widehat{v}_k(1-s).$$

Proof. The boundary condition $\widehat{\nu}_0(s) = \widehat{v}_0(1-s) = 0$ for all $s \in [0,1]$ gives us the induction case. Next, we prove that $\widehat{\nu}_k(s) = \widehat{v}_k(1-s)$ with the induction hypothesis

$$\widehat{\nu}_{k-1}(s) = \widehat{v}_{k-1}(1-s) \quad \text{ for all } s \in [0,1].$$

If we replace s with $1-s$ in the recursion (3.19), we obtain that

$$\widehat{v}_k(1-s) = \{1 - \widehat{h}_k(1-s)\}\widehat{v}_{k-1}(1-s) + \int_0^{\widehat{h}_k(1-s)} \{1 + \widehat{v}_{k-1}(1-s-w)\}\, dw.$$

By invoking the induction hypothesis, we can replace $\widehat{v}_{k-1}(1-s)$ with $\widehat{\nu}_{k-1}(s)$ and replace $\widehat{v}_{k-1}(1-s-w)$ with $\widehat{\nu}_{k-1}(s+w)$ to obtain that

$$\widehat{v}_k(1-s) = \{1 - \widehat{h}_k(1-s)\}\widehat{\nu}_{k-1}(s) + \int_0^{\widehat{h}_k(1-s)} \{1 + \widehat{\nu}_{k-1}(s+w)\}\, dw.$$

Then if we perform a change of variable $w = x - s$ for the integral on this last right-hand side, we obtain that

$$\widehat{v}_k(1-s) = \{1 - \widehat{h}_k(1-s)\}\widehat{\nu}_{k-1}(1-s) + \int_s^{\widehat{h}_k(1-s)+s} \{1 + \widehat{\nu}_{k-1}(x)\}\, dx.$$

Lastly, if we compare the two thresholds (3.16) and (3.17), and plug in the identity $\widehat{h}_k(1-s) = \widehat{\eta}_k(s) - s$ in (3.18), we obtain that

$$\widehat{v}_k(1-s) = \{1 - \widehat{\eta}_k(s) + s\}\}\widehat{\nu}_{k-1}(1-s) + \int_s^{\widehat{\eta}_k(s)} \{1 + \widehat{\nu}_{k-1}(x)\}\, dx. \tag{3.21}$$

By (3.20) we know that this last right-hand side is equal to $\widehat{\nu}_k(s)$, and hence $\widehat{\nu}_k(s) = \widehat{v}_k(1-s)$. As a result,

$$\widehat{\nu}_k(s) = \widehat{v}_k(1-s) \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1,$$

and the proof is complete. □

With this last proposition, we can apply our previous analysis of the dynamic and stochastic knapsack problem in Chapter 2 to prove the lower bound in Theorem 3.2. In fact, Proposition 3.2 tells us that the value function $\widehat{\nu}_k(s)$ of policy $\widehat{\pi}$ for the sequential increasing subsequence selection problem is equal to the value function $\widehat{v}_k(1-s)$ of the adaptive heuristic $\widehat{\pi}$ of Section 2.3 for the dynamic and stochastic knapsack problem with equal rewards. As we discussed in Section 2.4, the uniform distribution on the interval $[0,1]$ belongs to the typical class in Definition 2.1. Moreover, for this distribution, its associated constant in Lemma 2.2 is $M = 2$. Therefore, if we apply Theorem 2.1 to the value function $\widehat{v}_k(1-s)$, we would obtain that for all $s \in [0,1]$ and all $k \geqslant 1$,

$$\{2k(1-s)\}^{-1} - 2(\log(k) + 1) \leqslant \widehat{v}_k(1-s).$$

Then we obtain from Proposition 3.2 that

$$\{2k(1-s)\}^{-1} - 2(\log(k) + 1) \leqslant \widehat{\nu}_k(s) \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1,$$

completing the proof of the lower bound of (3.9) in Theorem 3.2.

We close this section with the comparison between the full-information version of the sequential monotone subsequence selection problem and that of the dynamic and stochastic knapsack problem. For a random sample $\{X_1, \ldots, X_n\}$, the full-information version of the sequential increasing subsequence selection problem is to find the longest increasing subsequence with length

$$L_n \equiv \max\left\{j : X_{i_1} \leqslant X_{i_2} \leqslant \cdots \leqslant X_{i_j} \text{ where } 1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant n\right\}. \tag{3.22}$$

This last solution is a sample-path upper bound for the sequential increasing subsequence selection problem. That is, for the optimal policy (3.2) of the sequential increasing subsequence selection

problem, one has that

$$L_n(\pi^*) \leqslant L_n \quad \text{almost surely.}$$

However, the upper bound $\mathbb{E}\left[L_n(\pi^*)\right] \leqslant (2n)^{1/2}$ in Theorem 3.1 does not apply to the full-information solution to the longest increasing subsequence selection problem. This does not come out very surprisingly because as we discussed in Section 3.2, such upper bound is intimately related to the fact that the optimal sequential policy is of a thresholding type. The solution (3.22) to the longest increasing subsequence selection problem, however, is not of a thresholding type due to its combinatorial nature. Moreover, for the sequential increasing subsequence selection problem, we know from Theorem 3.1 that

$$\mathbb{E}\left[L_n(\pi^*)\right] \sim (2n)^{1/2} \quad \text{as } n \to \infty.$$

While for the longest increasing subsequence selection problem, it has long been known that (Logan and Shepp, 1977; Veršik and Kerov, 1977)

$$\mathbb{E}\left[L_n\right] \sim 2(n)^{1/2} \quad \text{as } n \to \infty.$$

In other words, in the problems of monotone subsequence selection, a prophet with full information has a huge information advantage over the optimal sequential decision maker. This is in contrast with the dynamic and stochastic knapsack problem with equal rewards in which the online and offline solutions have the same first-order asymptotic performance. This is because both policies are given by a thresholding algorithm (Lemma 2.1). This feature is not present in the monotone subsequence selection problems in which only the online optimal algorithm is given by a sequence of threshold functions.

3.4 Connections and Observations

Policy $\widehat{\pi}$ is a simple adaptive online policy that selects a monotone increasing subsequence. It is then easy to see how it could be generalized to the sequential selection of a unimodal or d-modal subsequence considered by Arlotto and Steele (2011). For instance, in the unimodal case with n observations, one could set the turning time $\bar{n} = \lfloor n/2 \rfloor$ and run policy $\widehat{\pi}_{\bar{n}}$ to construct an increasing segment with the first $\bar{n}$ observations, and a decreasing version of $\widehat{\pi}_{n-\bar{n}}$ to obtain a decreasing segment over the next $n - \bar{n}$ observations. Theorem 3.1 then would immediately apply

to this construction, proving a $O(\log n)$-optimality gap for the sequential selection of unimodal and d-modal subsequences.

It is also reasonable to expect that policy $\widehat{\pi}$ could generalize to the sequential selection of coordinatewise increasing subsequences from a uniform random sample on the m-dimensional hypercube. Thus far, the best optimality-gap estimate comes from the work of Baryshnikov and Gnedin (2000) who use a non-adaptive policy to derive a $O(n^{1/(2m+2)})$ bound. The adaptive character of policy $\widehat{\pi}$ could be fruitful one more time and help establish a $O(\log n)$-optimality gap for this multidimensional problem.

In addition to the well-known bound $\mathbb{E}\,[L_n(\pi^*)] \leqslant (2n)^{1/2}$, Seksenbayev (2018) and Gnedin and Seksenbayev (2019) recently prove that the gap of this last upper bound has the asymptotics

$$g(n) = (2n)^{1/2} - \mathbb{E}\,[L_n(\pi^*)] \sim (12)^{-1} \log(n) \quad \text{ as } n \to \infty.$$

This last asymptotics tells us that our bound of Theorem 3.1 on the optimality gap is order tight. We also note that in the closely related problem in which the sequence $X_1, X_2, \ldots, X_n$ is given by a uniform random permutation of the integers $\{1, 2, \ldots, n\}$, Peng and Steele (2016) showed that the corresponding difference $g(n)$ is $O(\log n)$ as $n \to \infty$.

Several interesting questions remain open, however. Theorem 3.2 tells us that

$$\{2k(1-s)\}^{1/2} - \widehat{v}_k(s) \leqslant 2\{\log(k) + 1\} \qquad \text{for all } s \in [0,1] \text{ and all } k \geqslant 1,$$

and it would be worthwhile to understand how the right-hand side changes with the initial state value s. When $s = 0$ we have from Theorem 3.1 that

$$\mathbb{E}\,[L_n(\pi^*)] \leqslant \mathbb{E}\,[L_n(\widehat{\pi})] + 2\{\log(n) + 1\} \qquad \text{for all } n \geqslant 1,$$

but the actual expected performance of policy $\widehat{\pi}$ seems to be much tighter. Based on an extensive numerical analysis[1] (see also Figure 2.2), we conjecture that there is a constant $0 < c < \infty$ such that

$$\mathbb{E}\,[L_n(\pi^*)] \leqslant \mathbb{E}\,[L_n(\widehat{\pi})] + c \qquad \text{for all } n \geqslant 1.$$

Such conjecture is indeed a specialization of our early conjecture (2.41) of Section 2.8, and it would also imply that the functions $\widehat{\eta}_n, \widehat{\eta}_{n-1}, \ldots, \widehat{\eta}_1$ are remarkably close to their analogues that arise when implementing the optimal dynamic programming algorithm.

[1] We estimated numerically the value functions that solve the recursion (3.8) and its optimal counterpart on a discretized state space with a grid size of 10^{-5} and with k ranging from 1 to 10^4.

Chapter 4

Sequential Policies and the Distribution of Their Total Rewards in Dynamic and Stochastic Knapsack Problems

Originally introduced by Bellman (1957), stochastic dynamic programming is a technique for solving sequential decision problems in presence of uncertainty. Such problems can, in principle, be solved optimally by backward induction, but their solutions are rarely available in closed form, and they often suffer from the curse of dimensionality (Powell, 2011). As such, there is a pressing need for identifying approximate solutions with provable performance guarantees, and one common near-optimal criterion is that of asymptotic optimality (see, e.g., Coffman et al., 1987; Bruss and Robertson, 1991; Rhee and Talagrand, 1991; Gallego and van Ryzin, 1994, 1997; Talluri and van Ryzin, 1998). In this chapter, we study the reoptimized heuristic of Chapter 2 for the dynamic and stochastic knapsack problem with unitary rewards, and we show that the number of selections made by such policy has the same asymptotic variance and the same limiting distribution as that of the optimal dynamic programming policy. In contrast, we note that the number of selections made by another widely studied asymptotically optimal heuristic has larger asymptotic variance and a different limiting distribution. This contrast suggests that decision makers who are interested in assessing the higher-order effects of their near-optimal decisions should be careful with solely relying on asymptotic optimality.

To fix ideas and gain intuition, we now preview our main result by discussing a dynamic and stochastic knapsack problem with unitary rewards, and independent and identically distributed weights on $[0, 1]$. In such a problem, a decision maker (referred to as *she*) is given a knapsack with unit capacity, and she is presented with items arriving sequentially over n discrete time periods. As soon as an item arrives, its uniform weight $W_t, t \in [n] \equiv \{1, \ldots, n\}$ is revealed, and the decision maker has to decide whether to include the item in the knapsack or not. All decisions are terminal; they cannot be changed at a later time. The decision maker's objective is to maximize the

This chapter is written under the supervision of Prof. Alessandro Arlotto. The results presented here are also in a joint research paper Arlotto and Xie (2020b).

expected number of selections under the capacity constraint, and this can be achieved by formulating the problem as a dynamic program, or by proposing near-optimal heuristics. For instance, if the remaining knapsack capacity at the beginning of period t is x, then Chapter 2 proposes the *adaptive* (time-dependent) heuristic $\widehat{\pi}$ that selects the t-th arriving item if and only if its weight $W_t \leqslant \widehat{h}_{n-t+1}(x) \equiv \min\left\{x, \sqrt{\frac{2x}{n-t+1}}\right\}$. This decision rule then gives us a sequence of stopping times $0 \equiv \widehat{\tau}_0 < \widehat{\tau}_1 < \cdots < \widehat{\tau}_k$ such that

$$\widehat{\tau}_j = \min\left\{t : \widehat{\tau}_{j-1} < t \leqslant n \text{ and } W_t \leqslant \widehat{h}_{n-t+1}\left(1 - \sum_{i=1}^{j-1} W_{\widehat{\tau}_i}\right)\right\} \quad \text{for } j \geqslant 1, \tag{4.1}$$

and a total number of item selections given by

$$\widehat{N}_n = \max\left\{j \in \{0, 1, \ldots, n\} : \sum_{i=1}^{j} W_{\widehat{\tau}_i} \leqslant 1\right\}. \tag{4.2}$$

One can also choose alternative thresholds to make selection decisions. Bruss and Robertson (1991); Rhee and Talagrand (1991) suggest the *non-adaptive* policy $\widetilde{\pi}$ that selects the t-th arriving item if and only if its weight $W_t \leqslant \widetilde{h}_{n-t+1}(x) \equiv \min\left\{x, \sqrt{\frac{2}{n}}\right\}$, where—just as before—the variable x denotes the remaining knapsack capacity. One then has a sequence of stopping times analogous to (4.1) with the associated number of item selections $\widetilde{N}_n$.

Finally, one can also consider the thresholds $h_n^*, \ldots, h_1^*$ associated with the optimal dynamic programming policy π^* (Papastavrou et al., 1996) with its optimal number of item selections N_n^*. Here the three policies above play a key role because they all share the same first-order asymptotic performance.

In a sequence of papers (Samuels and Steele, 1981; Coffman et al., 1987; Rhee and Talagrand, 1991; Arlotto et al., 2018; Arlotto and Xie, 2020a), it has been shown that, all the three sequential policies have the same first-order performance with the closed-form expression

$$\mathbb{E}\left[\widetilde{N}_n\right] \sim \mathbb{E}\left[\widehat{N}_n\right] \sim \mathbb{E}\left[N_n^*\right] \sim \sqrt{2n} \quad \text{as } n \to \infty.$$

However, the second-order performances of the three policies are different. Specifically, Boshuizen and Kertz (1999) proves that the non-adaptive heuristic $\widetilde{\pi}$ has the variance asymptotics

$$\mathrm{Var}\left\{\widetilde{N}_n\right\} \sim \frac{2(\pi - 1)}{3\pi}\sqrt{2n} \quad \text{as } n \to \infty,$$

as well as the limiting behavior given as

$$\frac{\widetilde{N}_n - \sqrt{2n}}{\sqrt[4]{2n}} \Rightarrow Z_1 + (Z_2 \wedge 0) \quad \text{as } n \to \infty,$$

where Z_1 and Z_2 are centered normal random variables with covariance matrix $\begin{pmatrix} 1 & -1 \\ -1 & 4/3 \end{pmatrix}$.

On the other hand, we will discover as corollary of the main result in this chapter that the other two policies have the same variance asymptotics

$$\text{Var}\left\{\widehat{N}_n\right\} \sim \text{Var}\left\{N_n^*\right\} \sim \frac{1}{3}\sqrt{2n} \quad \text{as } n \to \infty,$$

which is smaller than that of the non-adaptive heuristic. In addition, with the same centering and scaling, our main result also implies the limit theorems

$$\frac{\widehat{N}_n - \sqrt{2n}}{\sqrt[4]{2n}} \Rightarrow N\left(0, \frac{1}{3}\right) \quad \text{and} \quad \frac{N_n^* - \sqrt{2n}}{\sqrt[4]{2n}} \Rightarrow N\left(0, \frac{1}{3}\right) \quad \text{as } n \to \infty.$$

Several observations are in order (see also Table 4.1). First, all three policies are first-order equivalent since they have the same first-order asymptotics $\sqrt{2n}$. Second, the adaptive heuristic $\widehat{\pi}$ shares the same variance asymptotics as the optimal dynamic programming policy π^*, while the non-adaptive heuristic $\widetilde{\pi}$ has an asymptotic variance that is about 36% larger. As such, while policy $\widetilde{\pi}$ provides the same asymptotic mean performance as policies $\widehat{\pi}$ and π^*, it does so with much higher risk. Lastly, the adaptive heuristic $\widehat{\pi}$ also shares the same limiting distribution as the optimal dynamic programming policy π^*, while the non-adaptive heuristic $\widetilde{\pi}$ has a different limiting distribution. In other words, the adaptivity of $\widehat{\pi}$ fully recovers the probabilistic behavior of the optimal dynamic programming policy π^*, going beyond asymptotic optimality.

As we will see in Section 4.1, such observations carry through from this special case to a class of weight distributions that satisfies certain regularity conditions. Specifically, we show that the adaptive heuristic $\widehat{\pi}$ is equivalent to the optimal dynamic programming policy π^*. It provides the same asymptotic first-order performance, the same asymptotic second-order performance, and the same limiting distribution. In general, our results suggest that, when faced with sequential decision problems, decision makers should be well informed about the potential risks that come with choosing which heuristic to implement.

Table 4.1: Asymptotic performance comparison as $n \to \infty$ among three policies.

	Non-adaptive heuristic $\widetilde{\pi}$	Adaptive heuristic $\widehat{\pi}$	Optimal Dynamic programming policy π^*
$\mathbb{E}[N_n]$	$\sqrt{2n}$	$\sqrt{2n}$	$\sqrt{2n}$
$\mathrm{Var}\{N_n\}$	$\frac{2(\pi-1)}{3\pi}\sqrt{2n}$	$\frac{1}{3}\sqrt{2n}$	$\frac{1}{3}\sqrt{2n}$
$\frac{N_n-\sqrt{2n}}{\sqrt[4]{2n}}$	$Z_1 + (Z_2 \wedge 0)$	$N\left(0, \frac{1}{3}\right)$	$N\left(0, \frac{1}{3}\right)$

Organization of the Chapter

In Section 4.1, we revisit the adaptive heuristic $\widehat{\pi}$ of Chapter 2 and state our main result. In Section 4.2 we establish the key Lemma 4.1 to control the derivative $\widehat{v}_k'(x)$ of the value function associated with the adaptive heuristic $\widehat{\pi}$. This key lemma is then used in Section 4.3 to obtain the variance asymptotics of the number of selections made by the adaptive heuristic $\widehat{\pi}$. With this last variance asymptotics and related estimations, we close our proof of the main result in Section 4.4. Lastly, we conclude this chapter with some closing remarks in Section 4.5.

4.1 Dynamic and Stochastic Knapsack Problem with Unitary Rewards: A Distributional Perspective

As we discussed in the literature review of Section 2.1, for the dynamic and stochastic knapsack problem and related formulations, much attention has been put to the analyses of policies in terms of their mean performances and worst-case performances. However, the distributional perspective of the problem has been much less touched in the literature. A notable exception is the case in which items arrive deterministically in each period with unit rewards, have uniform weights that are drawn on the interval $[0, 1]$, and the knapsack has unit capacity. For this case, the results of Arlotto et al. (2015) imply that the number of selections made by the optimal sequential policy, $N_n^*(1)$, has the variance asymptotics

$$\mathrm{Var}\{N_n^*(1)\} \sim \frac{1}{3}(2n)^{1/2} \quad \text{as } n \to \infty.$$

In addition, one also has the convergence in distribution

$$\frac{3^{1/2}\left\{N_n^*(1) - (2n)^{1/2}\right\}}{(2n)^{1/4}} \Rightarrow N(0,1) \quad \text{as } n \to \infty.$$

However, such limiting results are only available for this special case, and our goal is to generalize such results to a broader class of weight distributions. We note that the available limiting results of the special case is obtained through the equivalence between the sequential monotone subsequence selection problem and the dynamic and stochastic knapsack problem with equal rewards (see Section 3.3). While we need a different proof strategy for the general case.

4.1.1 The Adaptive Heuristic $\widehat{\pi}$ Revisited

Let us recall the dynamic and stochastic knapsack problem formulation of Chapter 2, and focus our attention to the case in which $p = r = 1$. That is, the items arrive deterministically in each period and have unitary rewards. The consumption function of (2.2) then becomes

$$\epsilon_k(x) = \sup\left\{\epsilon \in [0,\infty) : \int_0^{\epsilon} w\, dF(w) \leqslant \frac{x}{k}\right\} \quad \text{for all } k \geqslant 1 \text{ and all } x \in [0,c]. \tag{4.3}$$

This consumption function in turn gives us the adaptive heuristic of Section 2.3, given as $\widehat{\pi} = \{\widehat{h}_n, \ldots, \widehat{h}_1\}$, where for all $k \in [n]$,

$$\widehat{h}_k(x) = \min\{x, \epsilon_k(x)\} \quad \text{for all } x \in [0,c]. \tag{4.4}$$

Such adaptive heuristic uses time- and state-dependent thresholds for selection decisions. Specifically, for the t-th arriving item with weight realization W_t, if the current capacity level is x, then the policy $\widehat{\pi}$ includes this item into the knapsack if and only if $W_t \leqslant \widehat{h}_{n-t+1}(x)$. Thus, one can recursively define the stochastic process $\{X_t\}_{t\in[n]}$ of the remaining capacity under policy $\widehat{\pi}$ by setting $X_0 = c$ and for all $t \in [n]$

$$X_t = \begin{cases} X_{t-1} & \text{if } W_t > \widehat{h}_{n-t+1}(X_{t-1}) \\ X_{t-1} - W_t & \text{if } W_t \geqslant \widehat{h}_{n-t+1}(X_{t-1}). \end{cases} \tag{4.5}$$

The number of selections $\widehat{N}_n(c)$ under policy $\widehat{\pi}$ is then given by

$$\widehat{N}_n(c) = \sum_{t=1}^{n} \mathbb{1}\left\{W_t \leqslant \widehat{h}_{n-t+1}(X_{t-1})\right\}.$$

We note that such definition of the number of selections is equivalent to the definition through stopping times in (4.2), and those stopping times given in (4.1) are the times at which selections occur.

Besides the remaining capacity process $\{X_t\}_{t\in[n]}$ and the number of selections $\widehat{N}_n(c)$, the adaptive heuristic $\widehat{\pi} = \{\widehat{h}_n, \ldots, \widehat{h}_1\}$ has its associated value functions $\{\widehat{v}_k(x)\}_{k\in[n]}$. With $\widehat{v}_0(x) \equiv 0$, these value functions are recursively defined as

$$\widehat{v}_k(x) = (1 - F(\widehat{h}_k(x)))\widehat{v}_{k-1}(x) + \int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x-w)\}\, f(w)\, dw \quad \text{for all } k \in [n]. \tag{4.6}$$

The first-order performance of $\widehat{\pi}$ is then given by the identity $\widehat{v}_n(c) = \mathbb{E}\left[\widehat{N}_n(c)\right]$. Lastly, by taking derivative of this last recursion, we obtain the recursion of the derivative of the value function

$$\widehat{v}'_k(x) = \begin{cases} (1 - F(x))\, \widehat{v}'_{k-1}(x) + \int_0^x \widehat{v}'_{k-1}(x-w) f(w)\, dw + f(x)(1 - \widehat{v}_{k-1}(x)) & \text{if } \epsilon_k(x) > x \\ (1 - F(\epsilon_k(x)))\, \widehat{v}'_{k-1}(x) + \int_0^{\epsilon_k(x)} \widehat{v}'_{k-1}(x-w) f(w)\, dw & \\ + \frac{1}{k\epsilon_k(x)} \{1 + \widehat{v}_{k-1}(x - \epsilon_k(x)) - \widehat{v}_{k-1}(x)\} & \text{if } \epsilon_k(x) \leqslant x. \end{cases} \tag{4.7}$$

If the item-weight distribution F is continuous with density f, then one has the integral representation that for all $0 \leqslant x \leqslant k\mu \equiv k\mathbb{E}[W_1]$,

$$\int_0^{\epsilon_k(x)} w f(w)\, dw = \frac{x}{k}. \tag{4.8}$$

Two right-away observations are in order. First, for all $x \in [0, k\mu]$, one has that $0 \leqslant \frac{(k-1)x}{k} \leqslant (k-1)\mu$, so

$$\int_0^{\epsilon_{k-1}\left(\frac{(k-1)x}{k}\right)} w f(w)\, dw = \frac{x}{k}.$$

Thus, we find that

$$\epsilon_{k-1}\left(\frac{(k-1)x}{k}\right) = \epsilon_k(x) \quad \text{for all } x \in [0, k\mu]. \tag{4.9}$$

Second, when F is continuous with density f, implicit function theorem tells us that $\epsilon_k(x)$ is differentiable on $(0, k\mu)$, and its derivative is given by

$$\epsilon'_k(x) = \frac{1}{k\epsilon_k(x) f(\epsilon_k(x))} \quad \text{for all } x \in (0, k\mu). \tag{4.10}$$

This last derivative representation then gives us the concavity of $F^\alpha(\epsilon_k(x))$ for any power $\alpha \in [0, 1]$. Such property is gathered in the following proposition.

Proposition 4.1 (Concavity)**.** *Let $0 \leqslant \alpha \leqslant 1$ be fixed, then for any $k \geqslant 1$ and any $x \in [0, k\mu]$, the map $x \mapsto F^\alpha(\epsilon_k(x))$ is concave in x.*

Proof. Recall from (4.10) that for any $x \in [0, k\mu]$, we have

$$\epsilon_k'(x) = \frac{1}{k\epsilon_k(x) f(\epsilon_k(x))}.$$

As a result, for all $0 \leqslant \alpha \leqslant 1$ fixed, the derivative of $F^\alpha(\epsilon_k(x))$ is

$$\frac{d}{dx} F^\alpha(\epsilon_k(x)) = \frac{\alpha}{k\epsilon_k(x) F^{1-\alpha}(\epsilon_k(x))},$$

which is monotone decreasing in x. Therefore, the map $x \mapsto F^\alpha(\epsilon_k(x))$ is concave in x. □

We close this section with the martingale structure of a normalized remaining capacity process induced by $\widehat{\pi}$, as well as an easy implication of such martingale structure. Specifically, if we consider the stopping time

$$\eta \equiv \min\left\{t \in \{0, 1, \ldots, n\} : \epsilon_{n-t}(X_t) > X_t \text{ or } X_t > (n-t)\mu\right\}, \tag{4.11}$$

then with appropriate normalization, the remaining capacity process up until η forms a martingale. Such observation is made precise in the following proposition.

Proposition 4.2 (Martingale structure associated with the remaining capacity process)**.** *Let the stopping time η be defined as in* (4.11)*, then the normalized remaining capacity process up until η, i.e., $\{(n-t)^{-1} X_t : t = 0, 1, \ldots, \eta\}$ is a martingale. That is,*

$$\mathbb{E}\left[\frac{X_t}{n-t} \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{X_{t-1}}{n-t+1} \quad \textit{for all } t = 1, \ldots, \eta.$$

Proof. For all $t \in [n]$, the construction of $\widehat{\pi}$ gives us that

$$\mathbb{E}\left[X_t \mid \mathcal{F}_{t-1}\right] = X_{t-1} - \int_0^{\widehat{h}_{n-t+1}(X_{t-1})} w f(w)\, dw.$$

When $t \leqslant \eta$ in particular, the definition (4.11) tells us two more conditions. First, for all $1 \leqslant t \leqslant \eta$,

$$\widehat{h}_{n-t+1}(X_{t-1}) = \min\{\epsilon_{n-t+1}(X_{t-1}), X_{t-1}\} = \epsilon_{n-t+1}(X_{t-1}).$$

Second, for all $1 \leqslant t \leqslant \eta$, the remaining capacity always satisfies $X_{t-1} \leqslant (n-t+1)\mu$. This, together with (4.8), implies that

$$\int_0^{\epsilon_{n-t+1}(X_{t-1})} w f(w)\, dw = \frac{X_{t-1}}{n-t+1}.$$

Putting the last three identities altogether, we obtain that for all $1 \leqslant t \leqslant \eta$,

$$\mathbb{E}\left[X_t \mid \mathcal{F}_{t-1}\right] = \frac{(n-t)X_{t-1}}{n-t+1},$$

or equivalently,

$$\mathbb{E}\left[\frac{X_t}{n-t} \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{X_{t-1}}{n-t+1}.$$

Therefore, $\{(n-t)^{-1}X_t : t = 0, 1, \ldots, \eta\}$ is a martingale. □

This last martingale structure, together with Proposition 4.1, gives us the following series upper bound.

Proposition 4.3 (Series upper bound)**.** *Let the remaining capacity process $\{X_t : t = 0, 1, \ldots, n\}$ under the adaptive heuristic $\widehat{\pi}$ be defined as in* (4.5)*, and let the stopping time η be defined as in* (4.11)*. Then for any fixed $0 < \alpha \leqslant 1$, we have that for all $0 \leqslant t \leqslant \eta - 1$,*

$$\sum_{j=t+1}^{\eta} \mathbb{E}\left[\frac{F^{\alpha}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{1-\alpha}} \,\middle|\, \mathcal{F}_t\right] \leqslant \alpha^{-1}\left\{(n-t)F(\epsilon_{n-t}(X_t))\right\}^{\alpha} \qquad \textit{a.s.}$$

Proof. From Proposition 4.1 we know that $F^{\alpha}(\epsilon_k(x))$ is concave in x, so we can apply Jensen's inequality to obtain that for $j \geqslant 2$,

$$\mathbb{E}\left[F^{\alpha}(\epsilon_{n-j+1}(X_{j-1})) \mid \mathcal{F}_{j-2}\right] \leqslant F^{\alpha}(\epsilon_{n-j+1}(\mathbb{E}\left[X_{j-1} \mid \mathcal{F}_{j-2}\right])).$$

For any realization of the stopping time η, Proposition 4.2 tells us that for $2 \leqslant j \leqslant \eta$,

$$\mathbb{E}\left[\frac{X_{j-1}}{n-j+1} \,\middle|\, \mathcal{F}_{j-2}\right] = \frac{X_{j-2}}{n-j+2},$$

which in turn implies that

$$\mathbb{E}\left[F^{\alpha}(\epsilon_{n-j+1}(X_{j-1})) \mid \mathcal{F}_{j-2}\right] \leqslant F^{\alpha}\left(\epsilon_{n-j+1}\left(\frac{(n-j+1)X_{j-2}}{n-j+2}\right)\right) = F^{\alpha}(\epsilon_{n-j+2}(X_{j-2})),$$

where in the last equality we have used identity (4.9). With repeated applications of Jensen's inequality and the martingale structure, we finally obtain that for each summand of the series,

$$\mathbb{E}\left[F^{\alpha}(\epsilon_{n-j+1}(X_{j-1})) \mid \mathcal{F}_i\right] \leqslant F^{\alpha}(\epsilon_{n-i}(X_i)).$$

Therefore, the series is bounded from above by

$$\sum_{j=t+1}^{\eta} \mathbb{E}\left[\frac{F^{\alpha}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{1-\alpha}} \,\middle|\, \mathcal{F}_t\right] \leqslant F^{\alpha}(\epsilon_{n-i}(X_i)) \sum_{j=t+1}^{\tau} \frac{1}{(n-t+1)^{1-\alpha}}.$$

Applying the crude bound $\eta \leqslant n$ as well as the basic series inequality $\sum_{j=t+1}^{n}(n-j+1)^{\alpha-1} \leqslant \alpha^{-1}(n-t)^{\alpha}$, we obtain that

$$\sum_{j=t+1}^{\eta} \mathbb{E}\left[\frac{F^{\alpha}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{1-\alpha}} \,\middle|\, \mathcal{F}_t\right] \leqslant \alpha^{-1}\left\{(n-t)F(\epsilon_{n-t}(X_t))\right\}^{\alpha}.$$

Since the above argument holds for any realization of η, it holds almost surely. □

4.1.2 Main Result

When it comes to performance analysis of the dynamic and stochastic knapsack problems, regularity conditions are common practice in the literature. Since we are interested in the limiting regime in which the knapsack capacity is fixed while the number of arriving items grows, the behavior of the item-weight distribution F around zero is particularly important to the asymptotic analysis. We specify the regularity conditions for our analysis in the following definition.

Definition 4.1 (Regular distributions.). Any non-negative continuous distribution F with density f is said to be *regular* if it satisfies the following conditions on $(0, \bar{w})$ for some $\bar{w} > 0$.

(i) Bounded density.

$$0 < m_{\ell} \leqslant f(w) \leqslant m_u < \infty \quad \text{for all } w \in (0, \bar{w}). \tag{4.12}$$

(ii) Smoothness and existence of limit. The density f is differentiable and there exists constant $r \in (0, \infty)$ such that

$$\lim_{w \to 0^+} \frac{wf'(w)}{f(w)} = \frac{1}{r} - 1. \tag{4.13}$$

(iii) Convergence rate. In the neighborhood of zero,

$$\left| \frac{F(w)}{wf(w)} - r \right| \leqslant Mw^{3/4}, \quad \text{for all } w \in (0, \bar{w}). \tag{4.14}$$

Logarithmic Regret Bound

As a quick implication of Condition (ii), we have a lower bound on the first-order performance of the adaptive heuristic $\widehat{\pi}$. Specifically, from Condition (ii), one has that

$$(wf(w))' = f(w)\left(\frac{wf'(w)}{f(w)} + 1\right) \geqslant 0$$

in a neighborhood of zero, hence the function $wf(w)$ is monotone increasing. Moreover, by L'Hospital's rule and (4.13), we know that

$$\lim_{w\to 0^+} \frac{wF(w)}{\int_0^w F(y)\,dy} = \lim_{w\to 0^+} \frac{F(w) + wf(w)}{F(w)} = \frac{1+r}{r} > 1.$$

Then it follows from Theorem 2.1 and Lemma 2.2 that

$$nF(\epsilon_n(c)) - \mathbb{E}\left[\widehat{N}_n(c)\right] = O\left(\log n\right). \tag{4.15}$$

We state our main result in the following theorem.

Theorem 4.1 (Main result)**.** *Consider the dynamic and stochastic knapsack problem with unitary rewards, initial capacity c and regular item-weight distribution F as defined in Definition 4.1. Let the associated constant r be as in* (4.13)*, if we set*

$$\gamma(r) \equiv \frac{2r^2}{(1+r)(1+2r)},$$

then the number of selections $\widehat{N}_n(c)$ of the adaptive heuristic policy $\widehat{\pi}$ satisfies the convergence in distribution

$$\frac{\widehat{N}_n(c) - nF(\epsilon_n(c))}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} \Rightarrow N(0,1) \qquad \text{as } n \to \infty.$$

Moreover, if the distribution F satisfies the growth condition

$$\frac{nF(\epsilon_n(c))}{(\log n)^2} \to \infty \qquad \text{as } n \to \infty, \tag{4.16}$$

then the number of selections $N_n^*(c)$ under the optimal dynamic programming policy π^* satisfies the same convergence in distribution

$$\frac{N_n^*(c) - nF(\epsilon_n(c))}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} \Rightarrow N(0,1) \qquad \text{as } n \to \infty.$$

The item-weight distribution F plays a key role in the proof of our main result. Next, we collect the properties of F when it satisfies the regularity conditions in Definition 4.1.

4.1.3 Properties of the Regular Class

Definition 4.1 imposes regularity conditions on the item-weight distribution F. Such regularity conditions have several implications that are easy to use in the analysis. We summarize the implications below. The first implication is that of Condition (ii) in Definition 4.1.

Proposition 4.4 (Implication of Condition (ii), behavior near zero)**.** *Condition (ii) in Definition 4.1 of the regular class implies that there exists an interval $(0, \tilde{w})$ such that for all $w \in (0, \tilde{w})$, the two auxiliary functions*

$$T_1(w) \equiv wf(w), \quad T_2(w) \equiv \frac{\left(\int_0^w tf(t)\,dt\right)^2}{w^2 f(w)} \tag{4.17}$$

are monotone increasing; and

$$\max\left\{\frac{wf'(w)}{f(w)}, \frac{wf(w)}{F(w)}\right\} \leqslant M \tag{4.18}$$

for some constant M that depends only on F.

Proof. From (4.13) we know that there exists an interval $(0, \bar{w}_1)$ such that for all $w \in (0, \bar{w}_1)$,

$$\frac{wf'(w)}{f(w)} + 1 \geqslant 0,$$

which in turn implies that $T_1'(w) = wf'(w) + f(w) \geqslant 0$ and hence $T_1(w)$ is increasing in $w \in (0, \bar{w}_1)$.

On the other hand, the limit (4.13) also implies the following limit

$$\lim_{w\to 0^+}\left\{\frac{w^2 f(w)}{\int_0^w tf(t)\,dt} - \frac{wf'(w)}{2f(w)} - 1\right\} = \lim_{w\to 0^+}\frac{2wf(w) + w^2 f'(w)}{wf(w)} - a/2 - 1 = 1 + a/2 > 0.$$

As a result, there exists an interval $(0, \bar{w}_2)$ such that for any $w \in (0, \bar{w}_2)$,

$$\frac{w^2 f(w)}{\int_0^w tf(t)\,dt} - \frac{wf'(w)}{2f(w)} - 1 \geqslant 0.$$

Now if we compute the derivative of $T_2(w)$ in $w \in (0, \bar{w}_2)$, we would get

$$\begin{aligned} T_2'(w) =& \frac{2\int_0^w tf(t)\,dt \cdot w^3 f^2(w) - \left(\int_0^w tf(t)\,dt\right)^2 \left(w^2 f'(w) + 2wf(w)\right)}{w^4 f^2(w)} \\ =& \frac{2\left(\int_0^w tf(t)\,dt\right)^2}{w^3 f(w)} \left(\frac{w^2 f(w)}{\int_0^w tf(t)\,dt} - \frac{wf'(w)}{2f(w)} - 1 \right) \geqslant 0, \end{aligned}$$

so $T_2(w)$ is increasing in $w \in (0, \bar{w}_2)$.

As for the upper bound (4.18), if we set $\tilde{w} = \min\{\bar{w}_1, \bar{w}_2\}$ and consider the continuous function $wf'(w)/f(w)$, then the existence of the limit (4.13) tells us that the this continuous function defined on $w \in (0, \tilde{w}]$ can be extended to the compact interval $w \in [0, \tilde{w}]$. Since a continuous function on a compact interval is always bounded, we have that

$$\sup_{w \in [0, \bar{w}]} \frac{wf'(w)}{f(w)} = M < \infty,$$

and apparently the constant M depends only on F. The same argument applies to the other continuous function $wf(w)/F(w)$ as well. In fact, the existence of the following finite limit

$$\lim_{w \to 0+} \frac{wf(w)}{F(w)} = \lim_{w \to 0+} \frac{wf'(w) + f(w)}{f(w)} = a + 1 < \infty$$

implies that the domain of $\frac{wf(w)}{F(w)}$ can also be extended to the compact interval $w \in [0, \tilde{w}]$. Hence, we have that

$$\sup_{w \in [0, \bar{w}]} \frac{wf(w)}{F(w)} = M < \infty,$$

with M depending only on F. We now have all the claimed properties of regular distribution within $(0, \tilde{w})$. □

The second useful implication is that of Condition (iii) in Definition 4.1.

Proposition 4.5 (Implication of Condition (iii))**.** *Condition (iii) in the Definition 4.1 implies that there exist positive constants M and $\bar{w}$ such that*

$$\max\left\{ \left| \frac{\int_0^w F(y)\,dy}{wF(w)} - \frac{r}{1+r} \right|, \left| \frac{\int_0^w yF(y)\,dy}{w^2 F(w)} - \frac{r}{1+2r} \right| \right\} \leqslant Mw^{3/4} \quad \textit{for all } w \in (0, \bar{w}). \tag{4.19}$$

Proof. To prove (4.19), we start with the first maximand of the left-hand side. From equation (4.14) we know that for all $w \in (0, \bar{w})$,

$$\left| \frac{\int_0^w F(y)\,dy}{\int_0^w yf(y)\,dy} - r \right| \leqslant Mw^{3/4}.$$

If we multiply both sides with $(1+r)^{-1}\int_0^w yf(y)dy$, we obtain that

$$(1+r)^{-1}\left| \int_0^w F(y)\,dy - r\int_0^w yf(y)\,dy \right| \leqslant (1+r)^{-1}Mw^{3/4}\int_0^w yf(y)\,dy.$$

Note that with an integration by parts, the left-hand side of this last inequality can be re-written as

$$\begin{aligned}(1+r)^{-1}\left| \int_0^w F(y)\,dy - r\int_0^w yf(y)\,dy \right| &= (1+r)^{-1}\left| \int_0^w F(y)\,dy - r\left(wF(w) - \int_0^w F(y)\,dy\right) \right| \\ &= \left| \int_0^w F(y)\,dy - \frac{r}{1+r}wF(w) \right|,\end{aligned}$$

so we have

$$\left| \int_0^w F(y)\,dy - \frac{r}{1+r}wF(w) \right| \leqslant (1+r)^{-1}Mw^{3/4}\int_0^w yf(y)\,dy.$$

By dividing both sides with $wF(w)$, we finally obtain that

$$\left| \frac{\int_0^w F(y)\,dy}{wF(w)} - \frac{r}{1+r} \right| \leqslant \frac{\int_0^w yf(y)\,dy}{(1+r)wF(w)}Mw^{3/4} \leqslant Mw^{3/4},$$

where the last inequality is because $r > 0$ and $\int_0^w yf(y)\,dy \leqslant wF(w)$.

As for the second maximand in (4.19), we know from (4.14) that

$$\left| \frac{wF(w)}{w^2 f(w)} - r \right| \leqslant Mw^{3/4} \quad \text{for all } w \in (0, \bar{w}),$$

so we further have that for all $w \in (0, \bar{w})$,

$$\left| \frac{\int_0^w yF(y)\,dy}{\int_0^w y^2 f(y)\,dy} - r \right| \leqslant Mw^{3/4}.$$

Then we multiply both sides with $(1+2r)^{-1}\int_0^w y^2 f(y)\,dy$ and use integration by parts to the integral $\int_0^w y^2 f(y)\,dy$ on the left-hand side to obtain that

$$\left| \int_0^w yF(y)\,dy - \frac{r}{1+2r}w^2F(w) \right| \leqslant (1+2r)^{-1}Mw^{3/4}\int_0^w y^2 f(y)\,dy.$$

Finally, by dividing both sides with $w^2F(w)$, we obtain that

$$\left| \frac{\int_0^w yF(y)\,dy}{w^2F(w)} - \frac{r}{1+2r} \right| \leqslant \frac{\int_0^w y^2 f(y)\,dy}{(1+2r)w^2F(w)} Mw^{3/4} \leqslant Mw^{3/4},$$

completing the second half of (4.19). □

To take full use of the properties of the regular class, we define constant

$$K_0 \equiv \max\left\{ \left\lceil \frac{c}{\int_0^{\min\{\bar{w},\tilde{w}\}} wf(w)\,dw} \right\rceil, 3 \right\}, \tag{4.20}$$

and restrict ourselves into the problem instances with more than K_0 items. In fact, the choice of K_0 guarantees us that the consumption function $\epsilon_k(x)$ always satisfies $0 \leqslant \epsilon_k(x) \leqslant \min\{\bar{w}, \tilde{w}\}$ for all $x \in [0, c]$ whenever $k \geqslant K_0$.

The smoothness assumption of the item-weight distribution F implies quite rich analytical properties, which are heavily used in the sequel analysis. We close this section with these properties of regular class distributions.

Proposition 4.6 (Properties of regular class.)**.** *Let F be a given distribution that belongs to the regular class, then there exists constant M_F such that for all $x \in (0, c]$ and all $k \geqslant K_0$, the following properties hold.*

(i) Bounded ratios.

$$1 \leqslant \frac{k\epsilon_k(x)F(\epsilon_k(x))}{x} \leqslant M_F, \tag{4.21}$$

and

$$1 \leqslant \frac{k\epsilon_k^2(x)f(\epsilon_k(x))}{x} \leqslant M_F. \tag{4.22}$$

(ii) Difference in prophet upper bound. *For any $w \in [0, x]$,*

$$0 \leqslant kF(\epsilon_k(x)) - kF(\epsilon_k(x-w)) - \frac{w}{\epsilon_k(x)} \leqslant \frac{kF(\epsilon_k(x))w^2}{x^2}. \tag{4.23}$$

(iii) A Monotone Relation.

$$k^2\epsilon_k^2(x)f(\epsilon_k(x)) \leqslant (k+1)^2\epsilon_{k+1}^2(x)f(\epsilon_{k+1}(x)). \tag{4.24}$$

(iv) Taylor-expansion type bounds. *For any* $w \in [0, x/2]$,

$$\frac{\epsilon_k(x)}{\epsilon_k(x-w)} - 1 \leqslant \frac{w}{k\epsilon_k^2(x)f(\epsilon_k(x))} + M_F \frac{w^2}{x^2}. \tag{4.25}$$

Moreover, for any fixed $\alpha \in (0, 1/2)$, *there exists a constant* $M_{F,\alpha}$ *such that*

$$\frac{x^{1-\alpha}\epsilon_k^\alpha(x)}{(x-w)^{1-\alpha}\epsilon_k^\alpha(x-w)} - 1 \leqslant \left(\frac{1-\alpha}{x} + \frac{\alpha}{k\epsilon_k^2(x)f(\epsilon_k(x))}\right) w + M_{F,\alpha}\frac{w^2}{x^2}. \tag{4.26}$$

Proof. (i) The upper bound in (4.21) was proved in Lemma 2.2, while the lower bound in (4.21) is an easy consequence of (4.8)— if we apply integration by parts in (4.8) to get

$$\epsilon_k(x)F(\epsilon_k(x)) - \int_0^{\epsilon_k(x)} F(w)\, dw = \frac{x}{k},$$

and drop the positive integral, we obtain the inequality

$$\epsilon_k(x)F(\epsilon_k(x)) \geqslant \frac{x}{k},$$

which in turn implies the lower bound in (4.21).

As for the second inequality (4.22), we know that when $k \geqslant K_0$, the consumption function satisfies that $\epsilon_k(x) \leqslant \tilde{w}$ for all $x \in [0, c]$. Therefore, the monotonicity of $wf(w)$ from (4.17) implies that

$$\frac{x}{k} = \int_0^{\epsilon_k(x)} wf(w)\, dw \leqslant \int_0^{\epsilon_k(x)} \epsilon_k(x)f(\epsilon_k(x))\, dw = \epsilon_k^2(x)f(\epsilon_k(x)),$$

which is the left inequality of (4.22). On the other hand, with some algebra, we know that

$$\frac{k\epsilon_k^2(x)f(\epsilon_k(x))}{x} = \frac{k\epsilon_k(x)F(\epsilon_k(x))}{x}\frac{\epsilon_k(x)f(\epsilon_k(x))}{F(\epsilon_k(x))}.$$

Combining the first bound (4.21) and the bound (4.18) from Proposition 4.4, we get the right inequality of the second bound

$$\frac{k\epsilon_k^2(x)f(\epsilon_k(x))}{x} \leqslant M_F.$$

(ii) This was proved in Equation (2.28) of Proposition 2.2.

(iii) By the choice of K_0 and the definition of the consumption function, we have that $\epsilon_{k+1}(x) \leqslant \epsilon_k(x) \leqslant \tilde{w}$ for all $x \in [0, c]$. Then the monotone relation (4.24) is just an easy corollary of the

monotonicity of the function $T_2(w)$ in (4.17) from Proposition 4.4. In fact, since $T_2(w)$ is increasing in $w \in (0, \tilde{w}]$, we have that $T_2(\epsilon_{k+1}(x)) \leqslant T_2(\epsilon_k(x))$, which is

$$\frac{x/(k+1)}{(k+1)\epsilon_{k+1}^2(x) f(\epsilon_{k+1}(x))} \leqslant \frac{x/k}{k\epsilon_k^2(x) f(\epsilon_k(x))}.$$

This last inequality in turn implies that

$$k^2 \epsilon_k^2(x) f(\epsilon_k(x)) \leqslant (k+1)^2 \epsilon_{k+1}^2(x) f(\epsilon_{k+1}(x)),$$

as desired.

(iv) We first prove the first inequality (4.25), and the second inequality (4.26) follows a similar argument. The proof is basically Taylor expansion together with a verification of the boundedness of the second order derivative. For any $w \in [0, x/2]$, we first show that $\epsilon_k(x)/\epsilon_k(x-w) \leqslant 2$. In fact, $0 \leqslant w \leqslant x/2$ implies that

$$1 \leqslant \frac{x}{x-w} \leqslant 2. \tag{4.27}$$

From (4.8), we also know that

$$\int_{\epsilon_k(x/2)}^{\epsilon_k(x)} w f(w)\, dw = \frac{x}{2k} = \int_0^{\epsilon_k(x/2)} w f(w)\, dw. \tag{4.28}$$

The monotonicity of $wf(w)$ from Proposition 4.4 implies that

$$\int_{\epsilon_k(x/2)}^{\epsilon_k(x)} w f(w)\, dw \geqslant \epsilon_k(x/2) f(\epsilon_k(x/2)) \left\{\epsilon_k(x) - \epsilon_k(x/2)\right\},$$

$$\int_0^{\epsilon_k(x/2)} w f(w)\, dw \leqslant \epsilon_k(x/2) f(\epsilon_k(x/2)) \epsilon_k(x/2).$$

By connecting these two inequalities through (4.28) we obtain that

$$\epsilon_k(x) \leqslant 2\epsilon_k(x/2),$$

which in turn tells us that for all $w \in [0, x/2]$,

$$\frac{\epsilon_k(x)}{\epsilon_k(x-w)} \leqslant \frac{\epsilon_k(x)}{\epsilon_k(x/2)} \leqslant 2. \tag{4.29}$$

Next, we calculate the first two derivatives of $1/\epsilon_k(x)$, and, with some algebra, obtain that

$$\frac{d}{dx}\frac{1}{\epsilon_k(x)} = -\frac{1}{k\epsilon_k^3(x) f(\epsilon_k(x))},$$

$$\frac{d^2}{dx^2}\frac{1}{\epsilon_k(x)} = \frac{1}{k^2\epsilon_k^5(x) f^2(\epsilon_k(x))}\left(3 + \frac{\epsilon_k(x) f'(\epsilon_k(x))}{f(\epsilon_k(x))}\right).$$

Therefore for any $w \in [0, x/2]$, the second derivative satisfies that

$$\begin{aligned}
&x^2\epsilon_k^2(x) \cdot \frac{d^2}{dw^2}\frac{1}{\epsilon_k(x-w)} \\
=&\frac{x^2\epsilon_k(x)}{(x-w)^2\epsilon_k(x-w)} \frac{(x-w)^2}{k^2\epsilon_k^4(x-w)f^2(\epsilon_k(x-w))}\left(3 + \frac{\epsilon_k(x-w)f'(\epsilon_k(x-w))}{f(\epsilon_k(x-w))}\right) \leqslant M_F,
\end{aligned}$$

where we have used (4.22) for the boundedness of the second factor; (4.18) for the boundedness of the last factor; and the fact

$$\frac{x^2\epsilon_k(x)}{(x-w)^2\epsilon_k(x-w)} \leqslant 8$$

that we have just proved (a combination of (4.27) and (4.29)). As a result, the second order Taylor expansion upper bound for $1/\epsilon_k(x-w)$ at $w = 0$ reads

$$\frac{1}{\epsilon_k(x-w)} \leqslant \frac{1}{\epsilon_k(x)} + \frac{w}{k\epsilon_k^3(x)f(\epsilon_k(x))} + \frac{M_F w^2}{x^2\epsilon_k(x)},$$

which is equivalent to inequality (4.25) as we wanted.

On the other hand, we can expand the other function $\left\{(x-w)^{\alpha-1}\epsilon_k^{-\alpha}(x-w)\right\}$ at $w = 0$ as well. With some algebra and simplifications, we obtain the first two derivatives

$$\begin{aligned}
\frac{d}{dx}\left((x-w)^{\alpha-1}\epsilon_k^{-\alpha}(x-w)\right) =&(\alpha-1)x^{\alpha-2}\epsilon_k^{-\alpha}(x) - \alpha k^{-1}x^{\alpha-1}\epsilon_k^{-\alpha-2}(x)f^{-1}(\epsilon_k(x)), \\
\frac{d^2}{dx^2}\left((x-w)^{\alpha-1}\epsilon_k^{-\alpha}(x-w)\right) =&(1-\alpha)(2-\alpha)x^{\alpha-3}\epsilon_k^{-\alpha}(x) + 2\alpha(1-\alpha)k^{-1}x^{\alpha-2}\epsilon_k^{-\alpha-2}(x)f^{-1}(\epsilon_k(x)) \\
&+\alpha(\alpha+2)k^{-2}x^{\alpha-1}\epsilon_k^{-\alpha-4}(x)f^{-2}(\epsilon_k(x)) + \alpha k^{-2}x^{\alpha-1}\epsilon_k^{-\alpha-3}(x)f'(\epsilon_k(x))f^{-2}(\epsilon_k(x)).
\end{aligned}$$

Similar to what we did for the previous expansion, it is simple but tedious work to check that for any $w \in [0, x/2]$,

$$x^{3-\alpha}\epsilon_k^{\alpha}(x)\left\{\frac{d^2}{dw^2}\left((x-w)^{\alpha-1}\epsilon_k^{-\alpha}(x-w)\right)\right\} \leqslant M_{F,\alpha},$$

hence the second order Taylor expansion bound holds, that is, for any $w \in [0, x/w]$,

$$\begin{aligned}
(x-w)^{\alpha-1}\epsilon_k^{-\alpha}(x-w) \leqslant&x^{\alpha-1}\epsilon_k^{-\alpha}(x) + \left\{(1-\alpha)x^{\alpha-2}\epsilon_k^{-\alpha}(x) + \alpha k^{-1}x^{\alpha-1}\epsilon_k^{-\alpha-2}(x)f^{-1}(\epsilon_k(x))\right\} w \\
&+ \frac{M_{F,\alpha}w^2}{x^{11/4}\epsilon_k^{1/4}(x)},
\end{aligned}$$

which is equivalent to the second inequality (4.26) as we wanted. □

4.2 Non-asymptotic Derivative Bounds

The analysis of the value function derivative $\widehat{v}'_k(x)$ plays a crucial role in our proof of Theorem 4.1. In a nutshell, if the derivative $\widehat{v}'_k(x)$ can be tightly estimated, then the fluctuations along the sequential decision process can also be well controlled. Such important intermediate step is given in the following lemma.

Lemma 4.1 (Non-asymptotic derivative bounds.)**.** *If the item-weight distribution F belongs to the regular class in Definition 4.1, then there exists a constant m that depends only on F such that for all $k \geqslant K_0$, the derivative $\widehat{v}'_k(x)$ is bounded by*

$$\ell_k(m,x) \leqslant \widehat{v}'_{k-1}(x) \leqslant u_k(m,x) \quad \text{for all } x \in [0,c],$$

where the two functions are given by

$$u_k(m,x) = \frac{1}{\epsilon_k(x)} + \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)}, \quad \text{and} \quad \ell_k(m,x) = \frac{1}{\epsilon_k(x)} - \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)}.$$

The proof of Lemma 4.1 is based on an induction argument that consists of two parts. Essentially, we split up the domain $x \in [0,c]$ into two regions according to the ratio $\frac{x}{\epsilon_k(x)}$ and analyze the two regions separately. Specifically, we define two constants

$$M_0 \equiv \max\left\{16M_F, \frac{m_u}{2m_\ell}, 2\right\} \tag{4.30}$$

$$m_0 \equiv \max\left\{\frac{8(M_F+1)}{M_0^{1/4}}, \frac{m_u(M_FM_0+1)\left[M_0(1+K_0^{-1})\right]^{7/4}}{m_\ell}, \right. \tag{4.31}$$

$$\left. \left[M_0(1+K_0^{-1})\right]^{3/4} + \frac{m_u(M_FM_0+1)\left[M_0(1+K_0^{-1})\right]^{7/4}}{m_\ell}, 1\right\},$$

and consider for $k \geqslant K_0+1$ the following two regions

$$\Omega_k \equiv \{x \in [0,c] : x \leqslant M_0\epsilon_k(x)\}, \quad \text{and} \quad \Omega_k^c \equiv \{x \in [0,c] : x > M_0\epsilon_k(x)\}.$$

The first part of the induction argument for proving Lemma 4.1 is to control the derivative $\widehat{v}'_k(x)$ in the region Ω_k. We have the following proposition.

Proposition 4.7 (Derivative bounds when capacity is small)**.** *If the item weight distribution F belongs to the regular class, then whenever $k \geqslant K_0+1$ and $x \leqslant M_0\epsilon_{k-1}(x)$, we have the derivative*

bound

$$\ell_k(m,x) \leqslant \widehat{v}'_{k-1}(x) \leqslant u_k(m,x) \quad \textit{for all } x \in [0,c],$$

for all $m \geqslant m_0$ *defined in* (4.31).

Proof. Step 1. We start with the recursion of $\widehat{v}'_k(x)$ and prove by induction that

$$\max_{x\in\Omega_k} |\widehat{v}'_k(x)| \leqslant k(M_F M_0 + 1)m_u/2 \text{ for all } k \geqslant 1,$$

where $\Omega_k = \{x \in [0,c] : x \leqslant M_0\epsilon_k(x)\}$.

Before the induction proof, let us prepare two estimations that will be used later. First, from the integral representation of the consumption function (4.8) and the boundedness of the density f from (4.12), we know that

$$\frac{x}{k} = \int_0^{\epsilon_k(x)} wf(w)\,dw \leqslant m_u \int_0^{\epsilon_k(x)} w\,dw = \frac{m_u\epsilon_k^2(x)}{2},$$

which yields the inequality

$$\frac{1}{k\epsilon_k(x)} \leqslant \frac{m_u\epsilon_k(x)}{2x}. \tag{4.32}$$

The second estimation is that when $x \leqslant M_0\epsilon_{k-1}(x)$, from Theorem 2.1 we obtain that for all $y \in [0,x]$,

$$0 \leqslant \widehat{v}_{k-1}(y) \leqslant (k-1)F(\epsilon_{k-1}(y)) \leqslant (k-1)F(\epsilon_{k-1}(x)) = \frac{(k-1)\epsilon_{k-1}(x)F(\epsilon_{k-1}(x))}{x}\frac{x}{\epsilon_{k-1}(x)}.$$

From (4.21) we know that the first factor of this last right-hand side is bounded by M_F; and the second factor is bounded by M_0 since $x \leqslant M_0\epsilon_{k-1}(x)$. Hence we have that for all x such that $x \leqslant M_0\epsilon_{k-1}(x)$,

$$0 \leqslant \widehat{v}_{k-1}(y) \leqslant M_F M_0 \quad \text{for all } y \in [0,x]. \tag{4.33}$$

Now we are in the position to prove by induction the upper bound $|\widehat{v}'_k(x)| \leqslant k(M_F M_0 + 1)m_u/2$ for all $x \in \Omega_k = \{x \in [0,c] : x \leqslant M_0\epsilon_k(x)\}$. Recall the recursion of $\widehat{v}'_k(x)$ in (4.7). Since it has two different representations depending on the relation between $\epsilon_k(x)$ and x, we further divide the region $\Omega_{k-1} \equiv \{x \in [0,c] : x \leqslant M_0\epsilon_{k-1}(x)\}$ into two sub-regions: $0 \leqslant x < \epsilon_k(x)$ and $\epsilon_k(x) \leqslant x \leqslant M_0\epsilon_{k-1}(x)$.

On the one hand, if $0 \leqslant x < \epsilon_k(x)$, then we take absolute value on both sides of the first line of (4.7) and use triangle inequality to obtain that

$$\begin{aligned}\left|\, \widehat{v}_k'(x) \,\right| &\leqslant (1 - F(x)) \left|\, \widehat{v}_{k-1}'(x) \,\right| + \int_0^x \left|\, \widehat{v}_{k-1}'(x-w) \,\right| f(w)\, dw + f(x) \,|\, 1 - \widehat{v}_{k-1}(x) \,| \\ &\leqslant \max_{y \in [0,x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + f(x) \,|\, 1 - \widehat{v}_{k-1}(x) \,|\,.\end{aligned}$$

From (4.21) we know that

$$\widehat{v}_{k-1}(x) \leqslant (k-1)F(\epsilon_{k-1}(x)) \leqslant \frac{M_F x}{\epsilon_{k-1}(x)} \leqslant \frac{M_F x}{\epsilon_k(x)} \leqslant M_F,$$

so if we apply this last estimation on $\widehat{v}_{k-1}(x)$ and (4.12) to the inequality of $|\, \widehat{v}_k'(x) \,|$, we find the upper bound

$$\left|\, \widehat{v}_k'(x) \,\right| \leqslant \max_{y \in [0,x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + f(x) \,|\, 1 - \widehat{v}_{k-1}(x) \,| \leqslant \max_{y \in [0,x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + M_F m_u.$$

On the other hand, if $\epsilon_k(x) \leqslant x \leqslant M_0 \epsilon_{k-1}(x)$, then we take absolute value on both sides of the second line of (4.7) and use triangle inequality to obtain that

$$\begin{aligned}\left|\, \widehat{v}_k'(x) \,\right| &\leqslant (1 - F(\epsilon_k(x))) \left|\, \widehat{v}_{k-1}'(x) \,\right| + \int_0^{\epsilon_k(x)} \left|\, \widehat{v}_{k-1}'(x-w) \,\right| f(w)\, dw \\ &\quad + \frac{1}{k\epsilon_k(x)} \,|\, 1 + \widehat{v}_{k-1}(x - \epsilon_k(x)) - \widehat{v}_{k-1}(x) \,| \\ &\leqslant \max_{y \in [x - \epsilon_k(x), x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + \frac{1}{k\epsilon_k(x)} \,|\, 1 + \widehat{v}_{k-1}(x - \epsilon_k(x)) - \widehat{v}_{k-1}(x) \,|\,.\end{aligned}$$

Next, we apply (4.32) and (4.33) to this last right hand to obtain the upper bound

$$\left|\, \widehat{v}_k'(x) \,\right| \leqslant \max_{y \in [x - \epsilon_k(x), x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + (M_F M_0 + 1) m_u / 2.$$

Combining these last two inequalities on $|\, \widehat{v}_k'(x) \,|$, we obtain that for all $x \in \Omega_{k-1}$,

$$\begin{aligned}\left|\, \widehat{v}_k'(x) \,\right| &\leqslant \max_{y \in [x - \widehat{h}_k(x), x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + \max\{M_F m_u, (M_F M_0 + 1) m_u / 2\} \\ &= \max_{y \in [x - \widehat{h}_k(x), x]} \left|\, \widehat{v}_{k-1}'(y) \,\right| + (M_F M_0 + 1) m_u / 2. \qquad (4.34)\end{aligned}$$

Lastly, we define $x_k \equiv \sup\{x \in [0, c] : x \in \Omega_k\}$ and prove by induction that for all $k \geqslant 1$,

$$\max_{x \in \Omega_k} |\widehat{v}_k'(x)| \leqslant \max_{x \in [0, x_k]} |\widehat{v}_k'(x)| \leqslant k(M_F M_0 + 1) m_u / 2.$$

In fact, the nesting relation $\Omega_k \subset \Omega_{k-1}$ for all $k \geqslant 2$ tells us that $x_k \leqslant x_{k-1}$ for all $k \geqslant 2$. Since $\widehat{v}_1(x) = F(x)$ for all $x \in [0, c]$, we have from (4.12) that

$$\max_{x\in[0,x_1]} \left| \widehat{v}_1'(x) \right| \leqslant \max_{x\in[0,c]} \left| \widehat{v}_1'(x) \right| = \max_{x\in[0,c]} f(x) \leqslant m_u \leqslant (M_F M_0 + 1) m_u/2,$$

so the induction base is satisfied. Then, if $\max_{x\in[0,x_{k-1}]} |\widehat{v}_{k-1}'(x)| \leqslant (k-1)(M_F M_0 + 1) m_u/2$, we have from (4.34) that

$$\begin{aligned}
\max_{x\in[0,x_k]} \left| \widehat{v}_k'(x) \right| &\leqslant \max_{x\in[0,x_k]} \max_{y\in[\widehat{h}_k(x),x]} \left| \widehat{v}_{k-1}'(y) \right| + (M_F M_0 + 1) m_u/2 \\
&= \max_{x\in[0,x_k]} \left| \widehat{v}_{k-1}'(x) \right| + (M_F M_0 + 1) m_u/2 \\
&\leqslant \max_{x\in[0,x_{k-1}]} \left| \widehat{v}_{k-1}'(x) \right| + (M_F M_0 + 1) m_u/2 \\
&\leqslant (k-1)(M_F M_0 + 1) m_u/2 + (M_F M_0 + 1) m_u/2 = k(M_F M_0 + 1) m_u/2, \qquad (4.35)
\end{aligned}$$

just as we needed.

Step 2. As the second step, we quantitatively estimate the two auxiliary functions $u_k(m, x)$ and $\ell_k(m, x)$ when $k \geqslant K_0 + 1$ and $x \leqslant M_0 \epsilon_{k-1}(x)$. First we apply the integral representation (4.8) twice to get the identity

$$\int_{\epsilon_k(x)}^{\epsilon_{k-1}(x)} w f(w)\, dw = \frac{x}{k(k-1)} = \frac{1}{k-1} \int_0^{\epsilon_k(x)} w f(w)\, dw. \qquad (4.36)$$

The monotonicity of the integrand $wf(w)$ from (4.17) implies a lower bound on the right integral in (4.36)

$$\int_{\epsilon_k(x)}^{\epsilon_{k-1}(x)} w f(w)\, dw \geqslant (\epsilon_{k-1}(x) - \epsilon_k(x))\, \epsilon_k(x) f(\epsilon_k(x)),$$

as well as an upper bound on the left integral in (4.36)

$$\int_0^{\epsilon_k(x)} w f(w)\, dw \leqslant \epsilon_k^2(x) f(\epsilon_k(x)).$$

Taking these last two bounds back to (4.36) and simplify, we obtain the inequality

$$\epsilon_{k-1}(x) \leqslant \left(1 + \frac{1}{k-1}\right) \epsilon_k(x) \leqslant \left(1 + K_0^{-1}\right) \epsilon_k(x), \qquad (4.37)$$

for all $k \geqslant K_0 + 1$.

Next, the integral representation of the consumption function (4.8) and the boundedness of the density f in (4.12) imply that

$$\frac{x}{k} = \int_0^{\epsilon_k(x)} wf(w)\,dw \geqslant m_\ell \int_0^{\epsilon_k(x)} w\,dw = \frac{1}{2} m_\ell \epsilon_k^2(x),$$

which in turn gives us an upper bound on the consumption function $\epsilon_k(x) \leqslant \sqrt{2x/(km_\ell)}$. This upper bound together with (4.37) yields that when $x \leqslant M_0\epsilon_{k-1}(x)$, we have that

$$x \leqslant M_0\epsilon_{k-1}(x) \leqslant M_0(1+K_0^{-1})\epsilon_k(x) \leqslant M_0(1+K_0^{-1})\sqrt{\frac{2x}{km_\ell}}.$$

Solving this inequality with respect to x, we obtain that

$$x \leqslant \frac{2M_0^2(1-K_0^{-1})^2}{km_\ell},$$

and hence

$$x^{-3/4}\epsilon_k^{-1/4}(x) \geqslant x^{-3/4}\left(\sqrt{\frac{2x}{km_\ell}}\right)^{-1/4} \geqslant \frac{km_\ell}{2\left[M_0(1+K_0^{-1})\right]^{7/4}}. \tag{4.38}$$

As a result, we have the lower bound

$$u_k(m,x) = \frac{1}{\epsilon_k(x)} + \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)} \geqslant \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)} \geqslant \frac{kmm_\ell}{2\left[M_0(1+K_0^{-1})\right]^{7/4}}.$$

Since $m \geqslant m_0$, in particular the second maximand in (4.31), we have that

$$u_k(m,x) \geqslant k(M_F M_0 + 1)m_u/2,$$

which, together with (4.35), imply that $\widehat{v}'_{k-1}(x) \leqslant u_k(m,x)$.

On the other hand, when $x \leqslant M_0\epsilon_{k-1}(x)$, we have the upper bound

$$\ell_k(m,x) = \frac{1}{\epsilon_k(x)} - \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)} = -\frac{m-(x/\epsilon_k(x))^{3/4}}{x^{3/4}\epsilon_k^{1/4}(x)} \leqslant -\frac{m - M_0^{3/4}[\epsilon_{k-1}(x)/\epsilon_k(x)]^{3/4}}{x^{3/4}\epsilon_k^{1/4}(x)}.$$

An application of (4.37) further implies that

$$\ell_k(m,x) \leqslant -\frac{m-[M_0(1+K_0^{-1})]^{3/4}}{x^{3/4}\epsilon_k^{1/4}(x)}.$$

Therefore, when $m \geqslant m_0$, in particular the third maximand in (4.31), we have that

$$\ell_k(m,x) \leqslant -\frac{m_u(M_F M_0+1)\left[M_0(1+K_0^{-1})\right]^{7/4}}{m_\ell x^{3/4}\epsilon_k^{1/4}(x)}.$$

Another application of (4.38) yields that

$$\ell_k(m,x) \leqslant -k(M_F M_0+1)m_u/2,$$

and therefore by (4.35) we know that $\ell_k(m,x) \leqslant \widehat{v}'_{k-1}(x)$, and the proof is complete. □

The second part of the induction argument for proving Lemma 4.1 is to estimate the derivative $\widehat{v}'_k(x)$ in the other region $\Omega_k^c = [0,c]\backslash\Omega_k = \{x \in [0,c] : x > M_0\epsilon_k(x)\}$. This estimation is given in the following proposition.

Proposition 4.8 (Intermediate step for induction)**.** *Consider the two auxiliary functions*

$$u_k(m,x) = \frac{1}{\epsilon_k(x)} + \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)}, \quad \textit{and} \quad \ell_k(m,x) = \frac{1}{\epsilon_k(x)} - \frac{m}{x^{3/4}\epsilon_k^{1/4}(x)},$$

with $m \geqslant m_0$ where m_0 is defined in (4.31)*. If the item-weight distribution F belongs to the regular class, then for all $k \geqslant K_0$ and all $x \in \Omega_k^c \equiv \{x \in [0,c] : x > M_0\epsilon_k(x)\}$, we have the following two inequalities*

$$u_{k+1}(m,x) \geqslant [1-F(\epsilon_k(x))]\,u_k(m,x) + \int_0^{\epsilon_k(x)} u_k(m,x-w)f(w)\,dw \tag{4.39}$$
$$+ \frac{1}{k\epsilon_k(x)}\left[1 - \int_0^{\epsilon_k(x)} u_k(m,x-w)\,dw\right],$$

$$\ell_{k+1}(m,x) \leqslant [1-F(\epsilon_k(x))]\,\ell_k(m,x) + \int_0^{\epsilon_k(x)} \ell_k(m,x-w)f(w)\,dw \tag{4.40}$$
$$+ \frac{1}{k\epsilon_k(x)}\left[1 - \int_0^{\epsilon_k(x)} \ell_k(m,x-w)\,dw\right].$$

Proof. With some simple rearrangements, inequality (4.39) is equivalent to

$$u_{k+1}(m,x) - u_k(m,x) + F(\epsilon_k(x))u_k(m,x) - \int_0^{\epsilon_k(x)} u_k(m,x-w)\left\{f(w) - \frac{1}{k\epsilon_k(x)}\right\}dw - \frac{1}{k\epsilon_k(x)} \geqslant 0.$$

Next, we plug in the representation $u_k(m,x) = \epsilon_k^{-1}(x) + mx^{-3/4}\epsilon_k^{-1/4}(x)$ and collect all the terms that result from $\epsilon_k^{-1}(x)$ as well as the last term $-(k\epsilon_k(x))^{-1}$ into one group $G_k(x)$, and all the rest terms into another group $H_k(x)$ to obtain that

$$u_{k+1}(m,x) - u_k(m,x) + F(\epsilon_k(x))u_k(m,x) - \int_0^{\epsilon_k(x)} u_k(m,x-w)\left\{f(w) - \frac{1}{k\epsilon_k(x)}\right\} dw - \frac{1}{k\epsilon_k(x)}$$

$$= G_k(x) + mH_k(x),$$

where the first group of terms are collected in

$$G_k(x) \equiv \frac{1}{\epsilon_{k+1}(x)} - \frac{1}{\epsilon_k(x)} - \frac{1}{k\epsilon_k(x)}\left\{1 - \int_0^{\epsilon_k(x)} \frac{1}{\epsilon_k(x-w)}\, dw\right\} - \int_0^{\epsilon_k(x)} \left\{\frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)}\right\} f(w)\, dw,$$

and the second group of terms are collected in

$$H(k,x) \equiv \frac{1}{x^{3/4}\epsilon_{k+1}^{1/4}(x)} - \frac{1}{x^{3/4}\epsilon_k^{1/4}(x)} + \frac{1}{k\epsilon_k(x)} \int_0^{\epsilon_k(x)} \frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)}\, dw$$

$$- \int_0^{\epsilon_k(x)} \left\{\frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)} - \frac{1}{x^{3/4}\epsilon_k^{1/4}(x)}\right\} f(w)\, dw.$$

Therefore, the first inequality (4.39) is equivalent to $G_k(x) + mH_k(x) \geqslant 0$. Due to the symmetricity between the functions $\ell_k(m,x)$ and $u_k(m,x)$, similar algebra implies that the second inequality (4.40) is equivalent to $G_k(x) - mH_k(x) \leqslant 0$. Hence, the proof of the two inequalities (4.39) and (4.40) will complete once we can show that

$$m \cdot H_k(x) \geqslant |\, G_k(x)\,| \quad \text{for all } x \in \Omega_k^c.$$

The rest of the proof consists of three steps. The first step is to give an upper bound on $|\, G_k(x)\,|$; the second step is to give a positive lower bound on $H_k(x)$; and the last step is to show that the choice of $m \geqslant m_0$ guarantees that the lower bound on $mH_k(x)$ always dominates the upper bound on $|\, G_k(x)\,|$.

Step 1. We will show that $|\, G_k(x)\,| \leqslant (M_F + 1)(kx)^{-1}$.

Recall that the function $G_k(x)$ is defined as

$$G_k(x) = \left\{\frac{1}{\epsilon_{k+1}(x)} - \frac{1}{\epsilon_k(x)}\right\} - \frac{1}{k\epsilon_k(x)}\left\{1 - \int_0^{\epsilon_k(x)} \frac{1}{\epsilon_k(x-w)}\, dw\right\}$$

$$- \int_0^{\epsilon_k(x)} \left\{\frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)}\right\} f(w)\, dw$$

$$= I_G + II_G + III_G.$$

Towards the upper bound on $|\, G_k(x)\,|$, we start with relaxing the first piece through the integral representation

$$I_G = \frac{1}{\epsilon_{k+1}(x)} - \frac{1}{\epsilon_k(x)} = \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} \frac{dw}{w^2} = \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} \frac{wf(w)}{w^3 f(w)}\, dw.$$

The monotonicity of $wf(w)$ from (4.17) tells us that $w^3 f(w)$ is also monotone increasing, and hence in the right-most integral of this last equation, if we replace the denominator $w^3 f(w)$ of the integrand with its upper bound $\epsilon_k^3(x) f(\epsilon_k(x))$, we get a lower bound

$$I_G \geqslant \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} \frac{wf(w)}{\epsilon_k^3(x) f(\epsilon_k(x))}\, dw = \frac{x}{k(k+1)\epsilon_k^3(x) f(\epsilon_k(x))}, \tag{4.41}$$

where the last identity is due to the representation of the consumption function (4.8). On the other hand, we have another integral representation of the first piece I_G given by

$$I_G = \frac{1}{\epsilon_{k+1}(x)} - \frac{1}{\epsilon_k(x)} = \frac{\epsilon_k(x) - \epsilon_{k+1}(x)}{\epsilon_k(x)\epsilon_{k+1}(x)} = \frac{1}{\epsilon_k(x)\epsilon_{k+1}(x)} \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} 1\, dw.$$

In this last right-hand side, if we use the monotonicity of $wf(w)$ from (4.17) once again to replace the integrand 1 with its upper bound $wf(w)/(\epsilon_{k+1}(x) f(\epsilon_{k+1}(x)))$, we find

$$I_G \leqslant \frac{1}{\epsilon_k(x)\epsilon_{k+1}(x)} \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} \frac{wf(w)}{\epsilon_{k+1}(x) f(\epsilon_{k+1}(x))}\, dw$$

Then we apply in turn the consumption function representation (4.8) and the monotone relation (4.24) to get

$$I_G \leqslant \frac{x}{k(k+1)\epsilon_k(x)\epsilon_{k+1}^2(x) f(\epsilon_{k+1}(x))} \leqslant \frac{(k+1)x}{k^3 \epsilon_k^3(x) f(\epsilon_k(x))}. \tag{4.42}$$

Putting both the lower bound (4.41) and the upper bound (4.42) together, we have the two-sided bound

$$\frac{x}{k(k+1)\epsilon_k^3(x) f(\epsilon_k(x))} \leqslant I_G \leqslant \frac{(k+1)x}{k^3 \epsilon_k^3(x) f(\epsilon_k(x))}. \tag{4.43}$$

For the second piece II_G of the function $G_k(x)$, we know from (4.10) the closed-form integral representation

$$\int \frac{1}{\epsilon_k(w)}\, dw = kF(\epsilon_k(w)),$$

which in turn gives us the closed-form identity

$$II_G = -\frac{1}{k\epsilon_k(x)}\left\{1 - \int_0^{\epsilon_k(x)} \frac{1}{\epsilon_k(x-w)}\,dw\right\} = \frac{1}{k\epsilon_k(x)}\left\{kF(\epsilon_k(x)) - kF(\epsilon_k(x-\epsilon_k(x))) - 1\right\}.$$

Now, if we apply (4.23) with $w = \epsilon_k(x)$, we get the inequality

$$0 \leqslant kF(\epsilon_k(x)) - kF(\epsilon_k(x-\epsilon_k(x))) - 1 \leqslant \frac{k\epsilon_k^2(x)F(\epsilon_k(x))}{x^2},$$

which further gives us the two-sided inequality on the second piece

$$0 \leqslant II_G \leqslant \frac{\epsilon_k(x)F(\epsilon_k(x))}{x^2}. \tag{4.44}$$

For the third piece

$$III_G = -\int_0^{\epsilon_k(x)} \left\{\frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)}\right\} f(w)\,dw,$$

since $x \in \Omega_k^c$, we know that $\epsilon_k(x) \leqslant M_0^{-1}x \leqslant x/2$ as $M_0 \geqslant 2$, so (4.25) implies that for all $w \in [0, \epsilon_k(x)]$ the first factor of the integrand of III_G satisfies the two-sided bound

$$\frac{w}{k\epsilon^3(x)f(\epsilon_k(x))} \leqslant \frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)} \leqslant \frac{w}{k\epsilon^3(x)f(\epsilon_k(x))} + M_F \cdot \frac{w^2}{x^2\epsilon_k(x)}.$$

Then we multiply this last equation with $f(w)$, integrate over $w \in [0, \epsilon_k(x)]$ and use the representation (4.8) to get

$$\begin{aligned}\frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))} &\leqslant \int_0^{\epsilon_k(x)} \left\{\frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)}\right\} f(w)\,dw \\ &\leqslant \frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))} + \frac{M_F}{x^2\epsilon_k(x)}\int_0^{\epsilon_k(x)} w^2 f(w)\,dw.\end{aligned}$$

In the rightmost integral of this last inequality, if we replace the integrand $w^2 f(w)$ with its upper bound $\epsilon_k(x)wf(w)$ and use (4.8) once again, we obtain that

$$\frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))} \leqslant \int_0^{\epsilon_k(x)} \left\{\frac{1}{\epsilon_k(x-w)} - \frac{1}{\epsilon_k(x)}\right\} f(w)\,dw \leqslant \frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))} + \frac{M_F}{kx}.$$

By multiplying this last equation with (-1) to reverse the direction, we obtain that

$$-\frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))} - \frac{M_F}{kx} \leqslant III_G \leqslant -\frac{x}{k^2\epsilon^3(x)f(\epsilon_k(x))}. \tag{4.45}$$

Lastly, we combine all the three inequalities (4.43), (4.44) and (4.45) to get the final estimation

$$-\frac{x}{k^2(k+1)\epsilon_k^3(x)f(\epsilon_k(x))}-\frac{M_F}{kx}\leqslant G_k(x)\leqslant\frac{x}{k^3\epsilon_k^3(x)f(\epsilon_k(x))}+\frac{M_F}{kx}.$$

Taking absolute value on this last equation, we obtain that

$$|G_k(x)|\leqslant\frac{x}{k^3\epsilon_k^3(x)f(\epsilon_k(x))}+\frac{M_F}{kx}. \tag{4.46}$$

If we rewrite the first term in this last right-hand side as

$$\frac{x}{k^3\epsilon_k^3(x)f(\epsilon_k(x))}=\frac{1}{kx}\left(\frac{x}{k\epsilon_k^2(x)f(\epsilon_k(x))}\right)\frac{x}{k\epsilon_k(x)},$$

and apply (4.22) to bound the second factor from above by 1, and apply (4.21) to bound the third factor from above by $F(\epsilon_k(x))$, we get

$$\frac{x}{k^3\epsilon_k^3(x)f(\epsilon_k(x))}\leqslant\frac{F(\epsilon_k(x))}{kx}\leqslant\frac{1}{kx}.$$

Putting this back to (4.46), we get the final estimation

$$|G_k(x)|\leqslant\frac{M_F+1}{kx}. \tag{4.47}$$

Step 2. We will show that $H_k(x)\geqslant\{8kx^{3/4}\epsilon_k^{1/4}(x)\}^{-1}$.

Recall that the function $H_k(x)$ is defined as

$$\begin{aligned}H_k(x)=&\frac{1}{x^{3/4}\epsilon_{k+1}^{1/4}(x)}-\frac{1}{x^{3/4}\epsilon_k^{1/4}(x)}+\frac{1}{k\epsilon_k(x)}\int_0^{\epsilon_k(x)}\frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)}\,dw\\&-\int_0^{\epsilon_k(x)}\left\{\frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)}-\frac{1}{x^{3/4}\epsilon_k^{1/4}(x)}\right\}f(w)\,dw\\=&I_H+II_H+III_H.\end{aligned}$$

Similar to what we did to the function $G_k(x)$, we exploit the monotonicity of $wf(w)$ to estimation the first piece I_H. In fact, we have the integral representation

$$I_H=x^{-3/4}\epsilon_{k+1}^{-1/4}(x)-x^{-3/4}\epsilon_k^{-1/4}(x)=\frac{1}{4x^{3/4}}\int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)}w^{-5/4}\,dw.$$

The monotonicity of $wf(w)$ from (4.17) implies that $w^{9/4}f(w)$ is also increasing, so for all $w \in [\epsilon_{k+1}(x), \epsilon_k(x)]$,

$$w^{-5/4} \geqslant \frac{wf(w)}{\epsilon_k^{9/4}(x) f(\epsilon_k(x))},$$

and as a result,

$$I_H \geqslant \frac{1}{4x^{3/4}} \int_{\epsilon_{k+1}(x)}^{\epsilon_k(x)} \frac{wf(w)}{\epsilon_k^{9/4}(x) f(\epsilon_k(x))}\, dw = \frac{x^{1/4}}{4k(k+1)\epsilon_k^{9/4}(x) f(\epsilon_k(x))}, \tag{4.48}$$

where the last equality is due to (4.8).

The second piece

$$II_H = \frac{1}{k\epsilon_k(x)} \int_0^{\epsilon_k(x)} \frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)}\, dw$$

has a crude lower bound if we replace the integrand $((x-w)^{3/4}\epsilon_k^{1/4}(x-w))^{-1}$ with its lower bound $(x^{3/4}\epsilon_k^{1/4}(x))^{-1}$ for all $w \in [0, \epsilon_k(x)]$. That is,

$$II_H \geqslant \frac{1}{k\epsilon_k(x)} \int_0^{\epsilon_k(x)} \frac{1}{x^{3/4}\epsilon_k^{1/4}(x)}\, dw = \frac{1}{kx^{3/4}\epsilon_k^{1/4}(x)}. \tag{4.49}$$

For the third piece

$$III_H = -\int_0^{\epsilon_k(x)} \left((x-w)^{-3/4}\epsilon_k^{-1/4}(x-w) - x^{-3/4}\epsilon_k^{-1/4}(x) \right) f(w)\, dw,$$

we take $\alpha = 1/4$ in (4.26) to get the bound on the first factor of the integrand

$$(x-w)^{-3/4}\epsilon_k^{-1/4}(x-w) - x^{-3/4}\epsilon_k^{-1/4}(x) \leqslant \left(\frac{3}{4x^{7/4}\epsilon_k^{1/4}(x)} + \frac{1}{4kx^{3/4}\epsilon_k^{9/4}(x) f(\epsilon_k(x))} \right) w + \frac{M_F w^2}{x^{11/4}\epsilon_k^{1/4}(x)}.$$

As a result, if we multiply this last inequality with $(-f(w))$ and integrate over $w \in [0, \epsilon_k(x)]$, we obtain that

$$III_H \geqslant -\int_0^{\epsilon_k(x)} \left(\frac{3}{4x^{7/4}\epsilon_k^{1/4}(x)} + \frac{1}{4kx^{3/4}\epsilon_k^{9/4}(x) f(\epsilon_k(x))} \right) wf(w) + \frac{M_F w^2 f(w)}{x^{11/4}\epsilon_k^{1/4}(x)}\, dw.$$

In the last summand of the integrand of this last right-hand side, if we replace $w^2 f(w)$ with its upper bound $\epsilon_k(x) wf(w)$, we obtain a further lower bound

$$III_H \geqslant -\int_0^{\epsilon_k(x)} \left(\frac{3}{4x^{7/4}\epsilon_k^{1/4}(x)} + \frac{1}{4kx^{3/4}\epsilon_k^{9/4}(x) f(\epsilon_k(x))} + \frac{M_F \epsilon_k^{3/4}(x)}{x^{11/4}} \right) wf(w)\, dw.$$

Another application of (4.8) yields that

$$III_H \geqslant -\frac{3}{4kx^{3/4}\epsilon_k^{1/4}(x)} - \frac{x^{1/4}}{4k^2\epsilon_k^{9/4}(x)f(\epsilon_k(x))} - \frac{M_F\epsilon_k(x)}{kx^{7/4}}. \tag{4.50}$$

Combining all three inequalities (4.48), (4.49) and (4.50) together, we obtain the overall estimation

$$H_k(x) \geqslant \left(\frac{k}{4(k+1)} - \frac{1}{4}\right)\frac{x^{1/4}}{k^2\epsilon_k^{9/4}(x)f(\epsilon_k(x))} + \left(\frac{1}{4} - \frac{M_F\epsilon_k(x)}{x}\right)\frac{1}{kx^{3/4}\epsilon_k^{1/4}(x)}.$$

From the left inequality in (4.22), we know that

$$\frac{x^{1/4}}{k^2\epsilon_k^{9/4}(x)f(\epsilon_k(x))} \leqslant \frac{1}{kx^{3/4}\epsilon_k^{1/4}(x)},$$

and hence

$$H_k(x) \geqslant \left(\frac{k}{4(k+1)} - \frac{M_F\epsilon_k(x)}{x}\right)\frac{1}{kx^{3/4}\epsilon_k^{1/4}(x)}.$$

The domain $\Omega_k^c = \{x \in [0,c] : x > M_0\epsilon_k(x)\}$ of x yields the further lower bound

$$H_k(x) \geqslant \left(\frac{k}{4(k+1)} - \frac{M_F}{M_0}\right)\frac{1}{kx^{3/4}\epsilon_k^{1/4}(x)}.$$

The choice of M_0 from (4.30) and K_0 from (4.20) guarantees us that the first factor in this last right-hand side is at least $\frac{k}{4(k+1)} - \frac{M_F}{M_0} \geqslant 1/8$, and so

$$H(k,x) \geqslant \frac{1}{8kx^{3/4}\epsilon_k^{1/4}(x)}. \tag{4.51}$$

Step 3. We will show that $m \geqslant m_0$ implies that $|\,G_k(x)\,| \leqslant m \cdot H_k(x)$.

In fact, from the choice of the constant m_0, specifically the first maximand in (4.31), we know that when $m \geqslant m_0$,

$$\left(\frac{8(M_F+1)}{m}\right)^4 \leqslant \left(\frac{8(M_F+1)}{m_0}\right)^4 \leqslant M_0.$$

Since $x \in \Omega_k^c = \{x \in [0,c] : x > M_0\epsilon_k(x)\}$, we further have that

$$\frac{8(M_F+1)}{m} \leqslant \frac{x^{1/4}}{\epsilon_k^{1/4}(x)},$$

which implies that

$$\frac{M_F+1}{kx} \leqslant \frac{m}{8kx^{3/4}\epsilon_k^{1/4}(x)}.$$

This last inequality together with the two estimations (4.47) and (4.51) imply that

$$|\, G_k(x)\,| \leqslant m \cdot H_k(x),$$

just as we needed. The proof is complete. □

With the last two propositions, we are now ready to prove the key lemma of this section.

Proof of Lemma 4.1. With the constant K_0 defined in (4.20), we consider an auxiliary function

$$\delta(x) \equiv \left| \widehat{v}'_{K_0-1}(x) \cdot x^{3/4}\epsilon_{K_0}^{1/4}(x) - \left(\frac{x}{\epsilon_{K_0}(x)}\right)^{3/4} \right|, \quad x \in (0,c].$$

The limit $\lim_{x\to 0}\epsilon_{K_0}(x) = 0$ together with the boundedness of density f from (4.12) enable us to apply L'Hospital's rule with (4.10) to get the limit

$$\lim_{x\to 0^+} \frac{x}{\epsilon_{K_0}(x)} = \lim_{x\to 0^+} K_0\epsilon_{K_0}(x) f(\epsilon_{K_0}(x)) = 0.$$

As a result, the function $\delta(x)$ defined on $(0,c]$ can be extended to the entire interval $[0,c]$, hence this continuous function $\delta(x)$ attains its finite maximum over the compact domain $[0,c]$. Then we can define constant m as

$$m \equiv \max\left\{ \max_{x\in[0,c]} \delta(x), m_0 \right\} < \infty. \tag{4.52}$$

This last definition of m gives us the induction base, that is,

$$\ell_{K_0}(m,x) \leqslant \widehat{v}'_{K_0-1}(x) \leqslant u_{K_0}(m,x), \quad \text{for all } x \in [0,c].$$

Now we proceed to the induction step. Assume that

$$\ell_k(m,x) \leqslant \widehat{v}'_{k-1}(x) \leqslant u_k(m,x), \quad \text{for all } x \in [0,c]. \tag{4.53}$$

Recall the k-th partition of the entire domain $[0,c] = \Omega_k \cup \Omega_k^c$, where $\Omega_k = \{x \in [0,c] : x \leqslant M_0\epsilon_k(x)\}$. From Proposition 4.7 we know that for all $x \in \Omega_k$,

$$\ell_{k+1}(m,x) \leqslant \widehat{v}'_k(x) \leqslant u_{k+1}(m,x).$$

Then to complete the proof, it suffices to show that (4.53) implies that for all $x \in \Omega_k^c$,

$$\ell_{k+1}(m,x) \leqslant \widehat{v}_k'(x) \leqslant u_{k+1}(m,x).$$

The integral representation (4.8) together with the boundedness of the density f from (4.12) yield the chain relation

$$\frac{x}{k} = \int_0^{\epsilon_k(x)} w f(w)\, dw \leqslant m_u \int_0^{\epsilon_k(x)} w\, dw = \frac{m_u \epsilon_k^2(x)}{2},$$

which in turn implies that

$$\frac{1}{k\epsilon_k(x)} \leqslant \frac{m_u \epsilon_k(x)}{2x}.$$

When $x \in \Omega_k^c = \{x \in [0,c] : x > M_0 \epsilon_k(x)\}$, this last inequality gives us a further upper bound

$$\frac{1}{k\epsilon_k(x)} \leqslant \frac{m_u}{2M_0}.$$

This last upper bound together with the boundedness of the density f from (4.12) as well as the definition of the constant M_0 in (4.30) imply that for all $w \in [0, \epsilon_k(x)]$,

$$f(w) \geqslant m_\ell \geqslant \frac{m_u}{2M_0} \geqslant \frac{1}{k\epsilon_k(x)}. \tag{4.54}$$

Now, let us turn back to the recursion of the derivative $\widehat{v}_k'(x)$. For all $x \in \Omega_k^c$, we know that $x > M_0\epsilon_k(x) \geqslant \epsilon_k(x)$, so the recursion of $\widehat{v}_k'(x)$ is given by

$$\begin{aligned}
\widehat{v}_k'(x) &= (1 - F(\epsilon_k(x)))\, v_{k-1}'(x) + \int_0^{\epsilon_k(x)} v_{k-1}'(x-w) f(w)\, dw + \frac{1}{k\epsilon_k(x)} \left(1 - \int_0^{\epsilon_k(x)} v_{k-1}'(x-w)\, dw\right) \\
&= (1 - F(\epsilon_k(x)))\, v_{k-1}'(x) + \int_0^{\epsilon_k(x)} v_{k-1}'(x-w) \left(f(w) - \frac{1}{k\epsilon_k(x)}\right) dw + \frac{1}{k\epsilon_k(x)}.
\end{aligned}$$

Then by the right inequality of our induction assumption (4.53) and the relation (4.54), we know that

$$\begin{aligned}
\widehat{v}_k'(x) &\leqslant (1 - F(\epsilon_k(x)))\, u_k(m,x) + \int_0^{\epsilon_k(x)} u_k(m, x-w) \left(f(w) - \frac{1}{k\epsilon_k(x)}\right) dw + \frac{1}{k\epsilon_k(x)} \\
&= (1 - F(\epsilon_k(x)))\, u_k(m,x) + \int_0^{\epsilon_k(x)} u_k(m, x-w) f(w)\, dw + \frac{1}{k\epsilon_k(x)} \left[1 - \int_0^{\epsilon_k(x)} u_k(m, x-w)\, dw\right],
\end{aligned}$$

so we can apply (4.39) to this last right-hand side to obtain that

$$\widehat{v}_k'(x) \leqslant u_{k+1}(m,x).$$

Parallely, by the left inequality of our induction assumption (4.53) and the relation (4.54), we have the lower bound

$$\begin{aligned}\widehat{v}_k'(x) &\geqslant (1-F(\epsilon_k(x)))\,\ell_k(m,x) + \int_0^{\epsilon_k(x)} \ell_k(m,x-w)\left(f(w) - \frac{1}{k\epsilon_k(x)}\right) dw + \frac{1}{k\epsilon_k(x)} \\ &= (1-F(\epsilon_k(x)))\,\ell_k(m,x) + \int_0^{\epsilon_k(x)} \ell_k(m,x-w) f(w)\, dw + \frac{1}{k\epsilon_k(x)}\left[1 - \int_0^{\epsilon_k(x)} \ell_k(m,x-w)\, dw\right].\end{aligned}$$

Then we apply (4.40) to this last right-hand side to obtain that

$$\widehat{v}_k'(x) \geqslant \ell_{k+1}(m,x).$$

To sum up, the induction assumption (4.53) does imply the two-sided inequality

$$\ell_{k+1}(m,x) \leqslant \widehat{v}_k'(x) \leqslant u_{k+1}(m,x), \quad \text{for all } x \in \Omega_k^c,$$

and therefore the inequality

$$\ell_k(m,x) \leqslant \widehat{v}_{k-1}'(x) \leqslant u_k(m,x) \quad \text{for all } x \in [0,c]$$

hold for all $k \geqslant K_0$.

We conclude this section with a corollary of Lemma 4.1 which states that the difference in the value function $|\,\widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w)\,|$ within the interval $w \in [0, \widehat{h}_k(x)]$ is uniformly bounded by constant that is independent of k.

Corollary 4.1 (Bounded difference in value functions within selection interval)**.** *There exists a constant $M \equiv \max\{4M_F, (m+1)(M_F/2+1)\}$ that depends only on F, such that for all $k \geqslant 1$ and all $x \in [0,c]$,*

$$|\,\widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w)\,| \leqslant M \quad \textit{for all } w \in [0, \min\{x, \epsilon_k(x)\}].$$

Proof. Take any $k \geqslant 1$ and $x \in [0,c]$ fixed. There are only two possible scenarios: (i) $x \leqslant 2\epsilon_k(x)$ or (ii) $x > 2\epsilon_k(x)$.

Case (i). If $x \leqslant 2\epsilon_k(x)$.

We apply Theorem 2.1 to obtain that, as a feasible online policy, $\widehat{\pi}_n(c)$, its value function has upper bound for all $y \in [0, x]$,

$$\widehat{v}_{k-1}(y) \leqslant (k-1)F(\epsilon_{k-1}(y)) \leqslant (k-1)F(\epsilon_{k-1}(x)) = \frac{(k-1)\epsilon_{k-1}(x)F(\epsilon_{k-1}(x))}{x}\frac{x}{\epsilon_{k-1}(x)}.$$

If we apply (4.21) to the first factor of this last right-hand side, and note that $x \leqslant 2\epsilon_k(x) \leqslant 2\epsilon_{k-1}(x)$, we find that

$$\widehat{v}_{k-1}(y) \leqslant 2M_F \quad \text{for all } y \in [0, x].$$

As a result, we obtain that if $x \leqslant 2\epsilon_k(x)$, then for all $w \in [0, \min\{x, \epsilon_k(x)\}] \subset [0, x]$,

$$|\, \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w)\,| \leqslant 4M_F.$$

Case (ii). If $x > 2\epsilon_k(x)$.

In this case, for all $w \in [0, \epsilon_k(x)]$, we have the relation $x - w \geqslant x - \epsilon_k(x) > \epsilon_k(x) \geqslant \epsilon_k(x-w)$, so

$$\frac{1}{(x-w)^{3/4}\epsilon_k^{1/4}(x-w)} = \left(\frac{\epsilon_k(x-w)}{x-w}\right)^{3/4}\frac{1}{\epsilon_k(x-w)} \leqslant \frac{1}{\epsilon_k(x-w)}.$$

This last inequality together with Lemma 4.1 imply that for all $y \in [x - \epsilon_k(x), x]$,

$$|\, \widehat{v}'_{k-1}(y)\,| \leqslant \frac{1}{\epsilon_k(y)} + \frac{m}{y^{3/4}\epsilon_k^{1/4}(y)} \leqslant \frac{m+1}{\epsilon_k(y)}.$$

Therefore, by the fundamental theorem of calculus, we have that for all $w \in [0, \epsilon_k(x)]$,

$$\begin{aligned}
&|\, \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w)\,| \\
&= \left|\int_{x-w}^{x} \widehat{v}'_{k-1}(y)\,dy\right| \leqslant \int_{x-w}^{x} |\, \widehat{v}'_{k-1}(y)\,|\,dy \leqslant \int_{x-\epsilon_k(x)}^{x} |\, \widehat{v}'_{k-1}(y)\,|\,dy \\
&\leqslant \int_{x-\epsilon_k(x)}^{x} \frac{m+1}{\epsilon_k(y)}\,dy = (m+1)\left[kF(\epsilon_k(x)) - kF(\epsilon_k(x-\epsilon_k(x)))\right],
\end{aligned}$$

where in the last identity we have used $\int \epsilon_k^{-1}(y)\,dy = kF(\epsilon_k(y))$ due to (4.10). If we set the parameter w in (4.23) to $w = \epsilon_k(x)$, we obtain that

$$kF(\epsilon_k(x)) - kF(\epsilon_k(x-\epsilon_k(x))) \leqslant \frac{kF(\epsilon_k(x))\epsilon_k^2(x)}{x^2} + 1 = \frac{kF(\epsilon_k(x))\epsilon_k(x)}{x}\frac{\epsilon_k(x)}{x} + 1.$$

Since $x > 2\epsilon_k(x)$, if we apply (4.21) once again to bound the first factor in this last right-hand side, we find that

$$kF(\epsilon_k(x)) - kF(\epsilon_k(x-\epsilon_k(x))) \leqslant M_F/2 + 1.$$

Therefore, we obtain the final bound that for all $w \in [0, \epsilon_k(x)]$,

$$| \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w) | \leqslant (m+1)(M_F/2+1).$$

To sum up, in either case, the difference in the value function $| \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w) |$ is bounded by constant that depends only on F, so the proof is complete. □

4.3 Variance Asymptotics with Remainder Bounds

Establishing the asymptotics of $\mathrm{Var}\left\{\widehat{N}_n(c)\right\}$ as $n \to \infty$ is one of the most important steps towards proving the weak convergence in Theorem 4.1. Due to the inter-dependency generated by the sequential decisions, we use martingales to analyze such inter-dependency.

4.3.1 Doob's Martingale and Conditional Variance of Martingale Differences

With the natural filtration $\{\mathcal{F}_t : t = 0, 1, \ldots, n\}$, we define Doob's martingale as

$$M_t \equiv \mathbb{E}\left[\widehat{N}_n(c) \,\middle|\, \mathcal{F}_{t-1}\right], \quad \text{for } t \in [n].$$

Then we also have the associated martingale differences

$$d_t = M_t - M_{t-1}, \quad \text{for all } t \in [n].$$

Towards proving the martingale central limit theorem, it is an important intermediate step to tightly estimate the variance of the martingale differences. To this end, we consider the $\mathcal{F}_t$-measurable conditional variance

$$\omega_{n-t}(X_t) \equiv \mathrm{Var}\left\{ \sum_{j=t+1}^{n} d_j \,\middle|\, \mathcal{F}_t \right\}. \tag{4.55}$$

Then by the orthogonality of martingale differences, we also have the equivalent representation

$$\omega_{n-t}(X_t) = \sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_t\right].$$

If we denote G_t as the event of selecting the t-th item under policy $\widehat{\pi}$, that is,

$$G_t \equiv \left\{\omega : W_t(\omega) \in [0, \widehat{h}_{n-t+1}(X_{t-1}(\omega))]\right\},$$

then for each $t \in [n]$, we have the decomposition of the martingale difference $d_t = A_t + B_t$, where the two quantities are given by

$$A_t \equiv \widehat{v}_{n-t}(X_{t-1}) - \widehat{v}_{n-t+1}(X_{t-1}),$$
$$B_t \equiv \{1 + \widehat{v}_{n-t}(X_{t-1} - W_t) - \widehat{v}_{n-t}(X_{t-1})\} \mathbb{1}\{G_t\}.$$

By Corollary 4.1 we know that $\max\{\| A_t \|_\infty, \| B_t \|_\infty\} \leqslant M + 1$, so we have the uniform bound

$$\| d_t \|_\infty = \| A_t + B_t \|_\infty \leqslant 2(M+1) \quad \text{for all } t \in [n]. \tag{4.56}$$

The zero-mean property of martingale difference $\mathbb{E}[d_t \mid \mathcal{F}_{t-1}] = 0$ translates into $A_t + \mathbb{E}[B_t \mid \mathcal{F}_{t-1}] = 0$, and hence

$$\mathbb{E}\left[d_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[B_t^2 \mid \mathcal{F}_{t-1}\right] - A_t^2. \tag{4.57}$$

The two quantities in this last right-hand side both have integral representations. From the value function recursion (4.6), we know that

$$A_t = -\int_0^{\widehat{h}_{n-t+1}(X_{t-1})} \{1 + \widehat{v}_{n-t}(X_{t-1} - w) - \widehat{v}_{n-t}(X_{t-1})\} f(w)\, dw. \tag{4.58}$$

On the other hand, we have the integral representation of the conditional expectation

$$\mathbb{E}\left[B_t^2 \mid \mathcal{F}_{t-1}\right] = \int_0^{\widehat{h}_{n-t+1}(X_{t-1})} \{1 + \widehat{v}_{n-t}(X_{t-1} - w) - \widehat{v}_{n-t}(X_{t-1})\}^2 f(w)\, dw. \tag{4.59}$$

Summing up identity (4.57) from $j = t+1$ to n and taking one more conditional expectation with respect to $\mathcal{F}_t$, we find that

$$\omega_{n-t}(X_t) = \sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_t\right] = \sum_{j=t+1}^{n} \mathbb{E}\left[B_j^2 \mid \mathcal{F}_t\right] - \sum_{j=t+1}^{n} \mathbb{E}\left[A_j^2 \mid \mathcal{F}_t\right]. \tag{4.60}$$

As we will see shortly, the summation $\sum_{j=t+1}^{n} \mathbb{E}\left[A_j^2 \mid \mathcal{F}_t\right]$ in this last identity is negligible compared to the other summation $\sum_{j=t+1}^{n} \mathbb{E}\left[B_j^2 \mid \mathcal{F}_t\right]$, which drives most of the fluctuation of the selection process. Fortunately, with the help of Lemma 4.1, we are able to tightly approximate $\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right]$ by replacing the integrand of the right-hand side of (4.59),

$$\{1 + \widehat{v}_{n-t}(X_{t-1} - w) - \widehat{v}_{n-t}(X_{t-1})\}^2 \quad \text{for } w \in [0, \widehat{h}_{n-t+1}(X_{t-1})],$$

with the square of the linear counterpart

$$\left(1 - \frac{w}{\widehat{h}_{n-t+1}(X_{t-1})}\right)^2 \quad \text{for } w \in [0, \widehat{h}_{n-t+1}(X_{t-1})].$$

Such approximation is then used to estimate the conditional variance $\omega_{n-t}(\widehat{X}_t)$.

4.3.2 Conditional Variance Bounds

The conditional variance $\omega_{n-t}(X_t)$ is estimated in the following lemma with upper bounds on the remainders.

Lemma 4.2 (Conditional variance bounds)**.** *There exists constant M that depends only on F and c such that for all $t \in [n]$,*

$$|\,\omega_{n-t}(X_t) - \gamma(r)v_{n-t}(X_t)\,| \leqslant M\left\{[(n-t)F(\epsilon_{n-t}(X_t))]^{1/4} + \log(n-t) + 1\right\}. \tag{4.61}$$

The proof of this last lemma relies on an intermediate step that approximates the important quantity

$$\int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x-w) - \widehat{v}_{k-1}(x)\}^2 f(w)\,dw$$

by replacing the integrand with the corresponding linear function. The error of this approximation is well controlled with the help of Lemma 4.1, as stated in the following proposition.

Proposition 4.9 (Linear approximation of integrand)**.** *There exists a constant M that depends only on F, such that for all $k \geqslant K_0$ with K_0 defined as in* (4.20)*, we have that*

$$\left|\int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x-w) - \widehat{v}_{k-1}(x)\}^2 f(w)\,dw - \int_0^{\widehat{h}_k(x)} \left(1 - \frac{w}{\widehat{h}_k(x)}\right)^2 f(w)\,dw\right| \tag{4.62}$$

$$\leqslant M\left(\frac{F^{1/4}(\epsilon_k(x))}{k^{3/4}} + \frac{1}{k}\right).$$

Proof. The left-hand side of (4.62) has the simple upper bound

$$\left|\int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x-w) - \widehat{v}_{k-1}(x)\}^2 f(w)\,dw - \int_0^{\widehat{h}_k(x)} \left(1 - \frac{w}{\widehat{h}_k(x)}\right)^2 f(w)\,dw\right|$$

$$\leqslant \sup_{w \in [0, \widehat{h}_k(x)]} \left|2 + \widehat{v}_{k-1}(x-w) - \widehat{v}_{k-1}(x) - \frac{w}{\widehat{h}_k(x)}\right| \int_0^{\widehat{h}_k(x)} \left|\widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w) - \frac{w}{\widehat{h}_k(x)}\right| f(w)\,dw.$$

From Corollary 4.1 we have the uniform bound for all $x \in [0, c]$ and all $k \geqslant 1$,

$$\sup_{w \in [0, \widehat{h}_k(x)]} |\, \widehat{v}_{k-1}(x - w) - \widehat{v}_{k-1}(x) \,| \leqslant M. \tag{4.63}$$

On the other hand, we also have the simple bound on the linear function

$$1 \leqslant 2 - \frac{w}{\widehat{h}_k(x)} \leqslant 2, \quad \text{for all } w \in [0, \widehat{h}_k(x)].$$

Thus,

$$\sup_{w \in [0, \widehat{h}_k(x)]} \left| 2 + \widehat{v}_{k-1}(x - w) - \widehat{v}_{k-1}(x) - \frac{w}{\widehat{h}_k(x)} \right| \leqslant M + 2,$$

and as a result, the left-hand side of (4.62) is bounded by

$$\left| \int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x - w) - \widehat{v}_{k-1}(x)\}^2 f(w)\, dw - \int_0^{\widehat{h}_k(x)} \left(1 - \frac{w}{\widehat{h}_k(x)}\right)^2 f(w)\, dw \right|$$

$$\leqslant (M + 2) \int_0^{\widehat{h}_k(x)} \left| \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x - w) - \frac{w}{\widehat{h}_k(x)} \right| f(w)\, dw.$$

Next, we recall the definition of the threshold function $\widehat{h}_k(x) = \min\{x, \epsilon_k(x)\}$ from (4.4) and proceed the analysis for the only two possible scenarios: (i) $\widehat{h}_k(x) = x$ and (ii) $\widehat{h}_k(x) = \epsilon_k(x)$ separately.

Case (i) $\widehat{h}_k(x) = x \leqslant \epsilon_k(x)$. The uniform bound (4.63) implies that the integrand of this last right-hand side satisfies that

$$\sup_{w \in [0, \widehat{h}_k(x)]} \left| \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x - w) - \frac{w}{\widehat{h}_k(x)} \right|$$

$$\leqslant \sup_{w \in [0, \widehat{h}_k(x)]} |\, \widehat{v}_{k-1}(x - w) - \widehat{v}_{k-1}(x) \,| + \sup_{w \in [0, \widehat{h}_k(x)]} \left| \frac{w}{\widehat{h}_k(x)} \right| \leqslant M + 1.$$

As a result, the left-hand side of (4.62) is further bounded from above by

$$\left| \int_0^{\widehat{h}_k(x)} \{1 + \widehat{v}_{k-1}(x - w) - \widehat{v}_{k-1}(x)\}^2 f(w)\, dw - \int_0^{\widehat{h}_k(x)} \left(1 - \frac{w}{\widehat{h}_k(x)}\right)^2 f(w)\, dw \right|$$

$$\leqslant (M + 2)(M + 1) F(\widehat{h}_k(x)).$$

With an application of (4.21), we know that

$$F(\widehat{h}_k(x)) \leqslant F(\epsilon_k(x)) \leqslant \frac{Mx}{k\epsilon_k(x)} \leqslant \frac{M_F}{k},$$

since $x \leqslant \epsilon_k(x)$. As a result, if $\widehat{h}_k(x) = x \leqslant \epsilon_k(x)$, then the left-hand side of (4.62) is bounded by $M_F(M+2)(M+1)/k$ from above.

Case (ii) $\widehat{h}_k(x) = \epsilon_k(x) \leqslant x$. We apply Lemma 4.1 as well as the fundamental theorem of calculus to obtain that

$$\begin{aligned} &\left| \widehat{v}_{k-1}(x) - \widehat{v}_{k-1}(x-w) - \frac{w}{\widehat{h}_k(x)} \right| = \left| \int_{x-w}^{x} \widehat{v}'_{k-1}(y) - \frac{1}{\epsilon_k(x)}\, dy \right| \\ \leqslant\ & \int_{x-w}^{x} \frac{1}{\epsilon_k(y)} - \frac{1}{\epsilon_k(x)}\, dy + m \int_{x-w}^{x} \frac{1}{y^{3/4}\epsilon_k^{1/4}(y)}\, dy. \end{aligned} \tag{4.64}$$

The first integral in this last right-hand side has the closed-form representation

$$\int_{x-w}^{x} \frac{1}{\epsilon_k(y)} - \frac{1}{\epsilon_k(x)}\, dy = kF(\epsilon_k(x)) - kF(\epsilon_k(x-w)) - \frac{w}{\epsilon_k(x)},$$

which, by (4.23), is bounded from above by $kF(\epsilon_k(x))w^2/x^2$, that is,

$$\int_{x-w}^{x} \frac{1}{\epsilon_k(y)} - \frac{1}{\epsilon_k(x)}\, dy \leqslant \frac{kF(\epsilon_k(x))w^2}{x^2}. \tag{4.65}$$

The integrand of the second integral satisfies that

$$\frac{1}{y^{3/4}\epsilon_k^{1/4}(y)} = \frac{1}{4\epsilon_k(y)(kF(\epsilon_k(y)))^{3/4}} \cdot 4 \left(\frac{k\epsilon_k(y)F(\epsilon_k(y))}{y} \right)^{3/4}.$$

The first factor is the derivative of $(kF(\epsilon_k(y)))^{1/4}$, while the second factor is bounded by $4M_F^{3/4}$ due to (4.21). As a result, we have that,

$$\begin{aligned} &\int_{x-w}^{x} \frac{1}{y^{3/4}\epsilon_k^{1/4}(y)}\, dy \\ \leqslant& 4M_F^{3/4} \int_{x-w}^{x} d\left[(kF(\epsilon_k(y)))^{1/4} \right] = 4M_F^{3/4} \left\{ (kF(\epsilon_k(x)))^{1/4} - (kF(\epsilon_k(x-w)))^{1/4} \right\} \\ =& 4M^{3/4} \frac{kF(\epsilon_k(x)) - kF(\epsilon_k(x-w))}{\left\{ (kF(\epsilon_k(x)))^{1/4} + (kF(\epsilon_k(x-w)))^{1/4} \right\} \left\{ (kF(\epsilon_k(x)))^{1/2} + (kF(\epsilon_k(x-w)))^{1/2} \right\}} \\ \leqslant& 4M^{3/4} \frac{kF(\epsilon_k(x)) - kF(\epsilon_k(x-w))}{(kF(\epsilon_k(x)))^{3/4}}. \end{aligned}$$

If we apply (4.23) on the numerator of this last right-hand side, we obtain a further upper bound

$$\int_{x-w}^{x} \frac{1}{y^{3/4}\epsilon_k^{1/4}(y)}\,dy \leqslant 4M^{3/4}\frac{w/\epsilon_k(x)+kF(\epsilon_k(x))w^2/x^2}{(kF(\epsilon_k(x)))^{3/4}}.$$

By taking this last inequality and (4.65) back to (4.64), we obtain that

$$\begin{aligned}\left|\widehat{v}_{k-1}(x)-\widehat{v}_{k-1}(x-w)-\frac{w}{\widehat{h}_k(x)}\right| \leqslant & \frac{kF(\epsilon_k(x))w^2}{x^2}+4M^{3/4}\frac{w/\epsilon_k(x)+kF(\epsilon_k(x))w^2/x^2}{(kF(\epsilon_k(x)))^{3/4}} \\ = & 4M^{3/4}\frac{w/\epsilon_k(x)}{(kF(\epsilon_k(x)))^{3/4}}+\frac{kF(\epsilon_k(x))w^2}{x^2}\left\{1+\frac{4M^{3/4}}{(kF(\epsilon_k(x)))^{3/4}}\right\}.\end{aligned}$$

Integrating this last equation against $dF(w)$ over $w\in[0,\epsilon_k(x)]$ leads us to

$$\begin{aligned}&\int_0^{\epsilon_k(x)}\left|\widehat{v}_{k-1}(x)-\widehat{v}_{k-1}(x-w)-\frac{w}{\widehat{h}_k(x)}\right| f(w)\,dw \\ \leqslant & \frac{4M^{3/4}x}{k\epsilon_k(x)(kF(\epsilon_k(x)))^{3/4}}+\frac{\epsilon_k(x)F(\epsilon_k(x))}{x}\left\{1+\frac{4M^{3/4}}{(kF(\epsilon_k(x)))^{3/4}}\right\} \\ = & \frac{4M^{3/4}F^{1/4}(\epsilon_k(x))}{k^{3/4}}\left\{\frac{x}{k\epsilon_k(x)F(\epsilon_k(x))}+\frac{\epsilon_k(x)}{x}\right\}+\frac{\epsilon_k(x)F(\epsilon_k(x))}{x}.\end{aligned}$$

By (4.21), we know that the second summand is bounded by M_F/k. While for second factor of the first summand, we notice that $\epsilon_k(x)/x\leqslant 1$ when $\widehat{h}_k(x)=\epsilon_k(x)$, and apply (4.21) again to obtain that

$$\frac{x}{k\epsilon_k(x)F(\epsilon_k(x))}\leqslant 1,$$

so overall we have the inequality

$$\int_0^{\epsilon_k(x)}\left|\widehat{v}_{k-1}(x)-\widehat{v}_{k-1}(x-w)-\frac{w}{\widehat{h}_k(x)}\right| f(w)\,dw \leqslant \frac{8M^{3/4}F^{1/4}(\epsilon_k(x))}{k^{3/4}}+\frac{M_F}{k},$$

and the proof is complete. □

Now, with this last proposition, we are ready to prove the key variance asymptotics result of this section.

Proof. (Proof of Lemma 4.2.) Recall the decomposition of $\omega_{n-t}(X_t)$ from (4.60)

$$\omega_{n-t}(X_t)=\sum_{j=t+1}^{n}\mathbb{E}\left[d_j^2\mid\mathcal{F}_t\right]=\sum_{j=t+1}^{n}\mathbb{E}\left[B_j^2\mid\mathcal{F}_t\right]-\sum_{j=t+1}^{n}\mathbb{E}\left[A_j^2\mid\mathcal{F}_t\right],$$

and the two summands from (4.58) and (4.59) respectively

$$A_j^2 = \left(\int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \{1 + \widehat{v}_{n-j}(X_{j-1} - w) - \widehat{v}_{n-j}(X_{j-1})\} f(w)\, dw \right)^2$$

$$\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] = \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \{1 + \widehat{v}_{n-j}(X_{j-1} - w) - \widehat{v}_{n-j}(X_{j-1})\}^2 f(w)\, dw.$$

The remaining capacity process $\{X_t\}_{t=0}^n$ behaves nicely before the policy becomes greedy, so we split the time horizon $t = 0, 1, \ldots, n$ into two parts: before and after the stopping time η, which is the first time that the policy $\widehat{\pi}$ becomes greedy. Recall from (4.11) that the stopping time η is

$$\eta = \min\{t \in [n] : \epsilon_{n-t}(X_t) > X_t \text{ or } X_t > (n-t)\mu\}.$$

Conservative horizon $1 \leqslant j \leqslant \eta$. The heuristic policy $\widehat{\pi}$ uses the conservative threshold $\widehat{h}_{n-j+1}(X_{j-1}) = \epsilon_{n-j+1}(X_{j-1})$. By Corollary 4.1 we know that the integrand of A_j is bounded by M, so

$$A_j^2 \leqslant M^2 F^2(\widehat{h}_{n-j+1}(X_{j-1})),$$

and by (4.62) in Proposition 4.9, we know that

$$\left| \mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\, dw \right|$$

$$\leqslant M \left(\frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} + \frac{1}{n-j+1} \right).$$

Combine these last two estimations together, we obtain that

$$\left| \sum_{j=t+1}^{\eta} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=t+1}^{\eta} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\, dw \right|$$

$$\leqslant \sum_{j=t+1}^{\eta} \left\{ M \left(\frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} + \frac{1}{n-j+1} \right) + M^2 F^2(\widehat{h}_{n-j+1}(X_{j-1})) \right\}. \tag{4.66}$$

Since for $1 \leqslant j \leqslant \eta$, we have $\widehat{h}_{n-j+1}(X_{j-1}) = \epsilon_{n-j+1}(X_{j-1}) \leqslant X_{j-1}$, we know that

$$\frac{F^2(\widehat{h}_{n-j+1}(X_{j-1}))}{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))/(n-j+1)^{1/4}}$$

$$= \left(\frac{(n-j+1)\epsilon_{n-j+1}(X_{j-1}) F(\epsilon_{n-j+1}(X_{j-1}))}{X_{j-1}} \right)^{3/4} \left(\frac{F(\epsilon_{n-j+1}(X_{j-1}))}{\epsilon_{n-j+1}(X_{j-1})} \right)^{3/4} \left(X_{j-1}^{3/4} F^{1/4}(\epsilon_{n-j+1}(X_{j-1})) \right).$$

The first factor in this last right hand side is bounded from above by $M_F^{3/4}$ due to (4.21); the second factor is bounded from above by $m_u^{3/4}$ due to (4.12); the last factor is bounded by $c^{3/4}$ since $X_{j-1} \leqslant X_0 = c$ and the quantity $F(\epsilon_{n-j+1}(X_{j-1}))$ is always smaller than 1, so we obtain that

$$\frac{F^2(\widehat{h}_{n-j+1}(X_{j-1}))}{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))/(n-j+1)^{1/4}} \leqslant (M_F m_u c)^{3/4},$$

which in turn implies that

$$F^2(\widehat{h}_{n-j+1}(X_{j-1})) \leqslant (M_F m_u c)^{3/4} \frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}}.$$

Substituting this inequality to the right-hand side of (4.66) leads us to

$$\begin{aligned}
&\left| \sum_{j=t+1}^{\eta} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - A_j^2 \right) - \sum_{j=t+1}^{\eta} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \right| \\
\leqslant & \sum_{j=t+1}^{\eta} \left\{ (M + M^2 (M_F m_u c)^{3/4}) \frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} + \frac{M}{n-j+1} \right\} \\
\leqslant & (M + M^2 (M_F m_u c)^{3/4}) \sum_{j=t+1}^{\tau} \frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} + M(\log(n-t)+1) \\
\leqslant & (M + M^2 (M_F m_u c)^{3/4}) \sum_{j=t+1}^{n} \frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} + M(\log(n-t)+1).
\end{aligned}$$

Take one more expectation on this last inequality with respect to the filtration $\mathcal{F}_t$ and apply Proposition 4.3, we find that

$$\begin{aligned}
&\mathbb{E}\left[\left| \sum_{j=t+1}^{\eta} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - A_j^2 \right) - \sum_{j=t+1}^{\eta} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \right| \middle| \mathcal{F}_t \right] \\
\leqslant & 4(M + M^2 (M_F m_u c)^{3/4}) \left\{ (n-t) F(\epsilon_{n-t}(X_t)) \right\}^{1/4} + M(\log(n-t)+1) \qquad (4.67)
\end{aligned}$$

Greedy horizon $\eta + 1 \leqslant j \leqslant n$. By the definition of η, we have either $X_\eta > (n-\eta)\mu$ or $X_\eta < \epsilon_{n-\eta}(X_\eta)$.

If $X_\eta > (n-\eta)\mu$, then we have that $n - \eta < X_\eta \mu^{-1} \leqslant X_0 \mu^{-1} = c\mu^{-1}$. As a result,

$$0 \leqslant \sum_{j=\eta+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \leqslant \sum_{j=\eta+1}^{n} F(\widehat{h}_{n-j+1}(X_{j-1})) \leqslant n - \eta \leqslant c\mu^{-1}.$$

From the uniform boundedness of the martingale differences in (4.56), we also have the simple bound

$$\left\| \sum_{j=\eta+1}^{n} d_j^2 \right\|_{\infty} \leqslant 4(n-\eta)(M+1)^2 \leqslant 4c\mu^{-1}(M+1)^2,$$

so we find the almost sure bound

$$\left| \sum_{j=\eta+1}^{n} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - A_j^2 \right) - \sum_{j=\eta+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \right| \leqslant 4c\mu^{-1}(M+1)^2. \tag{4.68}$$

If $X_\eta < \epsilon_{n-\eta}(X_\eta)$, then we apply Corollary 4.1 on the integrand of $\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right]$ to obtain that

$$\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] \leqslant M^2 F(\widehat{h}_{n-j+1}(X_{j-1})).$$

Then if we sum up this last estimation from $j = \eta + 1$ to n, we would get

$$0 \leqslant \sum_{j=\eta+1}^{n} \mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - \sum_{j=\eta+1}^{n} A_j^2 \leqslant \sum_{j=\eta+1}^{n} \mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] \leqslant M^2 \sum_{j=\eta+1}^{n} F(\widehat{h}_{n-j+1}(X_{j-1})).$$

Combining this last estimation with the simple bound

$$0 \leqslant \sum_{j=\eta+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \leqslant \sum_{j=\eta+1}^{n} F(\widehat{h}_{n-j+1}(X_{j-1})),$$

we find that

$$\left| \sum_{j=\eta+1}^{n} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - A_j^2 \right) - \sum_{j=\eta+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \right|$$

$$\leqslant M^2 \sum_{j=\eta+1}^{n} F(\widehat{h}_{n-j+1}(X_{j-1})).$$

For any realization of η, if we take expectation of this last inequality conditioning on X_η, we obtain that

$$\mathbb{E}\left[\left| \sum_{j=\eta+1}^{n} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1} \right] - A_j^2 \right) - \sum_{j=\eta+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\, dw \right| \,\middle|\, \eta, X_\eta \right]$$

$$\leqslant M^2 \sum_{j=\eta+1}^{n} \mathbb{E}\left[F(\widehat{h}_{n-j+1}(X_{j-1})) \,\middle|\, X_\eta \right]. \tag{4.69}$$

The inequality $X_\eta < \epsilon_{n-\eta}(X_\eta)$ tells us that the summation on this last right-hand side, which equals $\widehat{v}_{n-\eta}(X_\eta)$, satisfies

$$\widehat{v}_{n-\eta}(X_\eta) \leqslant (n-\eta)F(\epsilon_{n-\eta}(X_\eta)) = \frac{(n-\eta)\epsilon_{n-\eta}(X_\eta)F(\epsilon_{n-\eta}(X_\eta))}{X_\eta}\frac{X_\eta}{\epsilon_{n-\eta}(X_\eta)} \leqslant M_F, \tag{4.70}$$

due to (4.21). As a result, if we substitute this last upper bound to (4.69) and take one more expectation with respect to $\mathcal{F}_t$, we find that

$$\mathbb{E}\left[\left|\sum_{j=\eta+1}^{n}\left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=\eta+1}^{n}\int_0^{\widehat{h}_{n-j+1}(X_{j-1})}\left\{1-\frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \middle| \mathcal{F}_t\right] \leqslant M^2 M_F. \tag{4.71}$$

To sum up, in either case of (i) $X_\eta > (n-\eta)\mu$ or (ii) $X_\eta < \epsilon_{n-\eta}(X_\eta)$, from (4.68) and (4.71) we find that

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{j=\eta+1}^{n}\left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=\eta+1}^{n}\int_0^{\widehat{h}_{n-j+1}(X_{j-1})}\left\{1-\frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \middle| \mathcal{F}_t\right] \\ &\leqslant \max\left\{M^2 M_F, 4c\mu^{-1}(M+1)^2\right\}. \end{aligned} \tag{4.72}$$

Combining conservative and greedy horizons. From the conditional variance decomposition

$$\omega_{n-t}(X_t) = \sum_{j=t+1}^{n}\mathbb{E}\left[d_j^2 \mid \mathcal{F}_t\right] = \sum_{j=t+1}^{n}\mathbb{E}\left[B_j^2 \mid \mathcal{F}_t\right] - \sum_{j=t+1}^{n}\mathbb{E}\left[A_j^2 \mid \mathcal{F}_t\right],$$

if we combine (4.67) and (4.72) together and apply triangle inequality, we would get

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{j=t+1}^{n}\left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=t+1}^{n}\int_0^{\widehat{h}_{n-j+1}(X_{j-1})}\left\{1-\frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \middle| \mathcal{F}_t\right] \\ \leqslant&\mathbb{E}\left[\left|\sum_{j=t+1}^{\eta}\left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=t+1}^{\eta}\int_0^{\widehat{h}_{n-j+1}(X_{j-1})}\left\{1-\frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \middle| \mathcal{F}_t\right] \\ &+\mathbb{E}\left[\left|\sum_{j=\eta+1}^{n}\left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=\eta+1}^{n}\int_0^{\widehat{h}_{n-j+1}(X_{j-1})}\left\{1-\frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \middle| \mathcal{F}_t\right] \\ \leqslant&\left(M + M^2(M_F m_u c)^{3/4}\right)\left\{(n-t)F(\epsilon_{n-t}(X_t))\right\}^{1/4} + M(\log(n-t)+1) + \max\left\{M^2 M_F, 4c\mu^{-1}(M+1)^2\right\}\end{aligned}$$

Note that we always have the relation

$$\left|\omega_{n-t}(X_t) - \sum_{j=t+1}^{n} \mathbb{E}\left[\int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right]\right|$$

$$\leqslant \mathbb{E}\left[\left|\sum_{j=t+1}^{n} \left(\mathbb{E}\left[B_j^2 \mid \mathcal{F}_{j-1}\right] - A_j^2\right) - \sum_{j=t+1}^{n} \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right| \,\middle|\, \mathcal{F}_t\right],$$

so the proof will complete once we can show that

$$\left|\gamma(r)\widehat{v}_{n-t}(X_t) - \sum_{j=t+1}^{n} \mathbb{E}\left[\int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right]\right| \tag{4.73}$$

$$\leqslant M\left\{\left[(n-t)F(\epsilon_{n-t}(X_t))\right]^{1/4} + 1\right\}.$$

Recall that the value function $\widehat{v}_{n-t}(X_t)$ satisfies that $\widehat{v}_{n-t}(X_t) = \sum_{j=t+1}^{n} \mathbb{E}\left[F(\widehat{h}_{n-j+1}(X_{j-1}))\right]$, so the left-hand side of this last inequality can be represented as

$$\left|\sum_{j=t+1}^{n} \mathbb{E}\left[\gamma(r)F(\widehat{h}_{n-j+1}(X_{j-1})) - \int_0^{\widehat{h}_{n-j+1}(X_{j-1})} \left\{1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})}\right\}^2 f(w)\,dw\right]\right|.$$

To analyze each summand in this last representation, we introduce the shorthand function $\eta_k(x) : [0, c] \to \mathbb{R}$ defined as

$$\eta_k(x) \equiv \gamma(r)F(\epsilon_k(x)) - \int_0^{\epsilon_k(x)} \left\{1 - \frac{w}{\epsilon_k(x)}\right\}^2 f(w)\,dw,$$

and consider the quantity $k^{3/4}\eta_k(x)/F^{1/4}(\epsilon_k(x))$. In fact, we have the identity

$$\frac{k^{3/4}\eta_k(x)}{F^{1/4}(\epsilon_k(x))}$$

$$= \frac{k^{3/4}\int_0^{\epsilon_k(x)} \left[\frac{2r^2}{(1+r)(1+2r)} - \left(1 - \frac{w}{\epsilon_k(x)}\right)^2\right] f(w)\,dw}{F^{1/4}(\epsilon_k(x))}$$

$$= \frac{2k^{3/4}\int_0^{\epsilon_k(x)} \left(\frac{w}{\epsilon_k(x)} - \frac{1}{1+r}\right) f(w)\,dw}{F^{1/4}(\epsilon_k(x))} + \frac{k^{3/4}\int_0^{\epsilon_k(x)} \left(\frac{1}{1+2r} - \frac{w^2}{\epsilon_k^2(x)}\right) f(w)\,dw}{F^{1/4}(\epsilon_k(x))}$$

$$= 2k^{3/4}F^{3/4}(\epsilon_k(x))\left(\frac{r}{1+r} - \frac{\int_0^{\epsilon_k(x)} F(w)\,dw}{\epsilon_k(x)F(\epsilon_k(x))}\right) + 2k^{3/4}F^{3/4}(\epsilon_k(x))\left(\frac{\int_0^{\epsilon_k(x)} wF(w)\,dw}{\epsilon_k^2(x)F(\epsilon_k(x))} - \frac{r}{1+2r}\right)$$

Now, if we take absolute value on both sides and apply triangle inequality, we obtain that

$$\frac{k^{3/4}\,|\,\eta_k(x)\,|}{F^{1/4}(\epsilon_k(x))} \leqslant 2k^{3/4}F^{3/4}(\epsilon_k(x)) \left\{ \left| \frac{r}{1+r} - \frac{\int_0^{\epsilon_k(x)} F(w)\,dw}{\epsilon_k(x)F(\epsilon_k(x))} \right| + \left| \frac{\int_0^{\epsilon_k(x)} wF(w)\,dw}{\epsilon_k^2(x)F(\epsilon_k(x))} - \frac{r}{1+2r} \right| \right\}.$$

By (4.19), we have the further bound

$$\frac{k^{3/4}\,|\,\eta_k(x)\,|}{F^{1/4}(\epsilon_k(x))} \leqslant 2M\left(k\epsilon_k(x)F(\epsilon_k(x))\right)^{3/4},$$

where this last right-hand side is bounded by $2M(M_F x)^{3/4}$ due to (4.21). As a result,

$$|\,\eta_k(x)\,| \leqslant 2M(M_F x)^{3/4}\frac{F^{1/4}(\epsilon_k(x))}{k^{3/4}} \leqslant 2M(M_F c)^{3/4}\frac{F^{1/4}(\epsilon_k(x))}{k^{3/4}}.$$

This last estimation on $\eta_k(x)$ tells us that

$$\left| \sum_{j=t+1}^{\eta} \mathbb{E}\left[\gamma(r)F(\widehat{h}_{n-j+1}(X_{j-1})) - \int_0^{\hat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\,dw \,\middle|\, \mathcal{F}_t \right] \right|$$

$$\leqslant \sum_{j=t+1}^{\eta} \mathbb{E}\left[\,|\,\eta_{n-j+1}(X_{j-1})\,|\,|\,\mathcal{F}_t\right] \leqslant 2M(M_F c)^{3/4} \sum_{j=t+1}^{\eta} \mathbb{E}\left[\frac{F^{1/4}(\epsilon_{n-j+1}(X_{j-1}))}{(n-j+1)^{3/4}} \,\middle|\, \mathcal{F}_t \right].$$

An application of Proposition 4.3 yields that this last right-hand side has a further upper bound $8M(M_F c)^{3/4}\left\{(n-t)F(\epsilon_{n-t}(X_t))\right\}^{1/4}$. Lastly, (4.73) will become evident once one notices that the uncertainties after the stopping time τ is negligible. That is, $\widehat{v}_{n-\tau}(X_\tau) \leqslant M_F$ a.s., and

$$\sum_{j=\eta+1}^{n} \mathbb{E}\left[\int_0^{\hat{h}_{n-j+1}(X_{j-1})} \left\{ 1 - \frac{w}{\widehat{h}_{n-j+1}(X_{j-1})} \right\}^2 f(w)\,dw \right] \leqslant \sum_{j=\eta+1}^{n} \mathbb{E}\left[F(\widehat{h}_{n-j+1}(X_{j-1})) \right]$$

$$\leqslant \max\{M_F, K_0\} \quad \text{a.s.},$$

due to (4.70) and the definition of η. □

4.4 Martingale Central Limit Theorem

Our proof of Theorem 4.1 relies on a basic version of the martingale central limit theorem which is contained in, e.g., McLeish (1974, Theorem 2.3).

Proposition 4.10 (Martingale central limit theorem)**.** *Let $\{Z_j\}_{j=1}^n$ be a martingale difference sequence with respect to the increasing σ-filed $\{\mathcal{F}_j\}_{j=1}^n$. If*

$$\max_{1\leqslant j\leqslant n} \|\, Z_j \,\|_\infty \to 0 \quad \text{as } n \to \infty, \tag{4.74}$$

and

$$\sum_{j=1}^{n} \mathbb{E}\left[Z_j^2 \mid \mathcal{F}_{j-1}\right] \xrightarrow{p} 1 \quad \text{as } n \to \infty, \tag{4.75}$$

then we have the convergence in distribution

$$\sum_{j=1}^{n} Z_j \Rightarrow N(0,1) \quad \text{as } n \to \infty.$$

In order to verify (4.75) in our context, we need the following proposition.

Proposition 4.11 (A second conditional variance bound)**.** *There exists a constant M that depends only on c and F such that for all $n \geqslant 1$,*

$$\operatorname{Var}\left\{\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]\right\} \leqslant M(nF(\epsilon_n(c)))\left(\left[nF(\epsilon_n(c))\right]^{1/2} + (\log n)^2 + 1\right).$$

Proof. To simplify notations, we let $V = \sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]$ and define another martingale as

$$V_t \equiv \mathbb{E}\left[V \mid \mathcal{F}_i\right] \quad \text{for all } t \in [n].$$

Then we define the associated martingale differences as $\Delta_t \equiv V_t - V_{t-1}$ for all $t \in [n]$. For each $1 \leqslant t \leqslant n$, since $\sum_{j=1}^{t} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]$ is $\mathcal{F}_{t-1}$-measurable, by the definition of Δ_t we have

$$\begin{aligned}
\Delta_t =& \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right] \,\middle|\, \mathcal{F}_t\right] - \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right] \,\middle|\, \mathcal{F}_{t-1}\right] \\
=& \mathbb{E}\left[\sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right] \,\middle|\, \mathcal{F}_t\right] - \mathbb{E}\left[\sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right] \,\middle|\, \mathcal{F}_{t-1}\right] \\
=& \sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_t\right] - \mathbb{E}\left[\sum_{j=t+1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_t\right] \,\middle|\, \mathcal{F}_{t-1}\right] \\
=& \omega_{n-t}(X_t) - \mathbb{E}\left[\omega_{n-t}(X_t) \mid \mathcal{F}_{t-1}\right].
\end{aligned}$$

Recall the event of selecting the t-th item

$$G_t \equiv \left\{\omega : W_t(\omega) \in [0, \widehat{h}_{n-t+1}(X_{t-1}(\omega))]\right\},$$

and condition on the occurrence of G_t, we have that

$$\omega_{n-t}(X_t) = \omega_{n-t}(X_{t-1}) + (\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))\mathbb{1}\left\{G_t\right\}.$$

Substitute this back to the representation of Δ_t, we find that

$$\Delta_t = (\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))\mathbb{1}\{G_t\} - \mathbb{E}\left[(\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right].$$

Hence by squaring this last identity and taking expectation with respect to the filtration $\mathcal{F}_{t-1}$, we obtain the conditional second moment bound

$$\begin{aligned}\mathbb{E}\left[\Delta_t^2 \mid \mathcal{F}_{t-1}\right] =& \mathbb{E}\left[(\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))^2\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right] \\ &- \mathbb{E}\left[(\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right]^2 \\ \leqslant& \mathbb{E}\left[(\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1}))^2\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right].\end{aligned}$$

By Lemma 4.2 and the triangle inequality, we know that

$$\begin{aligned}&|\,\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1})\,|\,\mathbb{1}\{G_t\} \\ \leqslant& |\,\omega_{n-t}(X_{t-1} - W_t) - \gamma(r)\widehat{v}_{n-t}(X_{t-1} - W_t)\,|\,\mathbb{1}\{G_t\} + |\,\omega_{n-t}(X_{t-1}) - \gamma(r)\widehat{v}_{n-t}(X_{t-1})\,| \\ &+ \gamma(r)\,|\,\widehat{v}_{n-t}(X_{t-1} - W_t) - \widehat{v}_{n-t}(X_{t-1})\,|\,\mathbb{1}\{G_t\} \\ \leqslant& 2M\left\{[(n-t)F(\epsilon_{n-t}(X_{t-1}))]^{1/4} + \log(n-t) + 1\right\} + \gamma(r)\,|\,\widehat{v}_{n-t}(X_{t-1} - W_t) - \widehat{v}_{n-t}(X_{t-1})\,|\,\mathbb{1}\{G_t\}.\end{aligned}$$

Corollary 4.1 tells us that the second summand in this last right-hand side is bounded by $M\gamma(r)$, so

$$|\,\omega_{n-t}(X_{t-1} - W_t) - \omega_{n-t}(X_{t-1})\,|\,\mathbb{1}\{G_t\} \leqslant 2M\left\{[(n-t)F(\epsilon_{n-t}(X_{t-1}))]^{1/4} + \log(n-t) + 1\right\} + M\gamma(r),$$

and hence

$$\mathbb{E}\left[\Delta_t^2 \mid \mathcal{F}_{t-1}\right] \leqslant M\left\{(nF(\epsilon_n(c)))^{1/2} + (\log n)^2 + 1\right\}\mathbb{E}\left[\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right]$$

As a result, we find that

$$\begin{aligned}&\mathrm{Var}\left\{\sum_{j=1}^{n}\mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]\right\} = \sum_{t=1}^{n}\mathbb{E}\left[\Delta_t^2\right] = \sum_{t=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\Delta_t^2\right] \mid \mathcal{F}_{t-1}\right] \\ \leqslant& M\left\{(nF(\epsilon_n(c)))^{1/2} + (\log n)^2 + 1\right\}\sum_{t=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\{G_t\} \mid \mathcal{F}_{t-1}\right]\right] \\ =& M\left\{(nF(\epsilon_n(c)))^{1/2} + (\log n)^2 + 1\right\}\mathbb{E}\left[\widehat{N}_n(c)\right] \leqslant M\left(nF(\epsilon_n(c))\right)\left\{(nF(\epsilon_n(c)))^{1/2} + (\log n)^2 + 1\right\},\end{aligned}$$

since $\mathbb{E}\left[\widehat{N}_n(c)\right] \leqslant nF(\epsilon_n(c))$. The proof is complete. □

Now, we have all the ingredients to prove Theorem 4.1.

Proof. (Proof of Theorem 4.1.) If we define the normalized martingale differences as

$$Z_j \equiv \frac{d_j}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} \quad \text{for all } 1 \leqslant j \leqslant n,$$

then the uniform boundedness (4.56) of the martingale differences together with the growth condition (4.16) imply the negligibility condition (4.74).

With the growth condition (4.16), the conditional variance bound (4.61) tells us that

$$\sum_{j=1}^{n} \mathbb{E}\left[d_j^2\right] = \omega_n(X_0) \sim \gamma(r)nF(\epsilon_n(c)) \quad \text{as } n \to \infty.$$

On the other hand, we know from Proposition 4.11 that

$$\mathrm{Var}\left\{\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]\right\} \sim o\left(\left[nF(\epsilon_n(c))\right]^2\right) \quad \text{as } n \to \infty.$$

Hence if we combine these last two asymptotics and apply Chebyshev's inequality, we obtain that for any $\delta > 0$ fixed,

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{j=1}^{n} \mathbb{E}\left[Z_j^2 \mid \mathcal{F}_{j-1}\right] - 1\right| > \delta\right\} =& \mathbb{P}\left\{\left|\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right] - \gamma(r)nF(\epsilon_n(c))\right| > \gamma(r)nF(\epsilon_n(c))\delta\right\} \\ \leqslant& \frac{\mathrm{Var}\left\{\sum_{j=1}^{n} \mathbb{E}\left[d_j^2 \mid \mathcal{F}_{j-1}\right]\right\}}{\left[nF(\epsilon_n(c))\right]^2 \gamma(r)^2\delta^2} \to 0 \quad \text{as } n \to 0.\end{aligned}$$

Therefore, the convergence in probability condition (4.75) is satisfied and so by Proposition 4.10 we have the convergence in distribution

$$\sum_{j=1}^{n} Z_j = \sum_{j=1}^{n} \frac{d_j}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} = \frac{\widehat{N}_n(c) - \mathbb{E}\left[\widehat{N}_n(c)\right]}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} \Rightarrow N(0,1) \quad \text{as } n \to \infty.$$

Finally, the identity $\mathbb{E}\left[\widehat{N}_n(c)\right] = nF(\epsilon_n(c)) + O(\log n)$ from (4.15) together with the growth condition (4.16) lead us to the weak law

$$\frac{\widehat{N}_n(c) - nF(\epsilon_n(c))}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} \Rightarrow N(0,1) \quad \text{as } n \to \infty.$$

For the weak convergence of $N_n^*(c)$, we apply Theorem 2.1 to obtain that

$$\frac{\mathbb{E}\left[\left|\, N_n^*(c) - \widehat{N}_n(c) \,\right|\right]}{\sqrt{\gamma(r)nF(\epsilon_n(c))}} = \frac{O(\log n)}{\sqrt{\gamma(r)nF(\epsilon_n(c))}}.$$

The growth condition (4.16) then implies that the last right-hand side vanishes as $n \to \infty$. As a result,

$$\frac{N_n^*(c) - \widehat{N}_n(c)}{\sqrt{\gamma(r) n F(\epsilon_n(c))}} \to 0 \quad \text{in } L^1,$$

and hence we have that

$$\frac{N_n^*(c) - nF(\epsilon_n(c))}{\sqrt{\gamma(r) n F(\epsilon_n(c))}} \Rightarrow N(0,1) \quad \text{as } n \to \infty,$$

just as we wanted. □

4.5 Concluding Remarks

In this chapter, we study the dynamic and stochastic knapsack problem with unitary rewards and revisit the the adaptive heuristic policy $\widehat{\pi}$ of Chapter 2. We show that the number of selections made by such heuristic shares the same variance asymptotics and limiting distribution with that made the optimal dynamic programming policy. However, as we have noted, different asymptotically optimal policies may have different higher moments as well as different limiting behaviors. Such differences may be overlooked if asymptotic optimality is the only criterion to judge different policies.

Looking forward, we believe that the analysis of higher moments and the limiting distribution of the total reward collected by different policies is important when studying sequential decision problems. The analysis in this chapter serves as a first step towards this direction, and other related problems could potential by studied in a similar spirit. In addition to the mean-performance and worst-case-performance analyses, the complementary analysis of the higher moments and limiting behaviors of the total rewards collected by different policies would provide valuable and comprehensive managerial insights to decision makers.

Chapter 5

Data-driven Monitoring the Implementation of Policies across Many Markov Decision Problems

Markov decision problem (or MDP), with both finite and infinite horizon, is a commonly used framework for modeling sequential decision problems under uncertainty, and have been extensively studied in the literature. Numerous applications of MDP theory can be found in finance, operations management, operations research, robotics and telecommunication, just to name a few (see, e.g. White, 1993; Puterman, 1994; Bertsekas, 2005). In this chapter, we focus our attention to many discrete time finite-horizon MDPs faced in parallel by self-interested decision makers, where the exogenous uncertainties faced by each decision maker have arbitrary and unknown correlation structure. Our goal is to simultaneously monitor whether each of the decision makers is implementing her designated policy, with only the knowledge of the marginal distributions of the exogenous uncertainty but not the correlation structure. The key feature of our monitoring is that the number of decision makers, p, is allowed to be much larger than the number of periods, n. We also provide data-driven procedure to construct simultaneous confidence intervals for the total rewards collected by each decision maker, with theoretical asymptotic validity.

In a typical finite horizon MDP with n discrete time periods, the mean-optimal policy π^* is the policy that delivers the maximal expected total reward. That is,

$$\pi^* = \arg \max_{\pi \in \Pi(n)} \mathbb{E}\left[R_n(\pi)\right],$$

where $\Pi(n)$ is the class of all non-anticipating sequential policies. It is well-known that under mild conditions, there exists a Markovian deterministic policy that is optimal. Such optimal policy π^* is often defined through Bellman equation with the associated value function given by

$$v_n^* = \mathbb{E}\left[R_n(\pi^*)\right] = \sup_{\pi \in \Pi(n)} \mathbb{E}\left[R_n(\pi)\right],$$

However, the total reward $R_n(\pi^*)$ collected by the optimal policy π^* is *ex ante* random, and its expected value v_n^* alone does not tell us about where $R_n(\pi^*)$ would actually land. The same

This chapter is written under the supervision of Prof. Alexandre Belloni. The results presented here are also in a joint research paper Belloni and Xie (2020).

randomness exists for the total reward collected by other near-optimal policies as well. When there are p correlated MDPs faced in parallel by self-interested decision makers labeled by $j \in [p] \equiv \{1, 2, \ldots, p\}$, the vector of total rewards $\boldsymbol{R}_n \in \mathbb{R}^p$ each of them collects is also random, and it undergoes fluctuations and correlations that are generated both endogenously (sequential decisions in response to uncertainty realizations) and exogenously (correlation of the exogenous uncertainty). Without knowing the correlation structure of the exogenous uncertainty, but only knowing their marginal distributions instead, we are interested in testing the *null hypothesis*

$$H_0 \equiv \{ \text{ every decision maker } j \in [p] \text{ is implementing her designated policy } \pi_j \} . \tag{5.1}$$

That is, we are interested in simultaneously monitoring many individual decision makers with the observation of the rewards they collect. We are also interested in constructing simultaneous confidence intervals for the total rewards $\boldsymbol{R}_n \in \mathbb{R}^p$ collected by each individual decision maker, with asymptotic validity.

For the case of one MDP, there are rich literature on limit theorems for the total reward collected by the optimal policy. Although the majority work along this line of research has focused on the infinite-horizon formulation in which the optimal policy is stationary, a notable exception is Arlotto and Steele (2016) who focuses on a finite-horizon formulation. They generalize the work of Dobrushin (1956a,b) on additive functionals of non-stationary Markov chains, and show that, in the finite-horizon MDP setting, under proper condition on the Dobrushin's contraction coefficient, the total reward collected by the optimal policy satisfies a central limit theorem (CLT) as the number of time periods, n, goes to infinity.

When CLTs are available, the above-mentioned hypothesis testing and the construction of confidence intervals are just simple exercises. However, when the "dimension" p becomes large, the CLTs start to fail, even for partial sums of independent and identically distributed (i.i.d.) random variables. In fact, for i.i.d. zero-mean random variables $Y_1, \ldots, Y_n \in \mathbb{R}^p$ with $\mathbb{E}[Y_1 Y_1'] = I_p$ (p-by-p identity matrix), and their Gaussian counterparts $\widetilde{Y}_i \sim N(0, I_p)$ for $i \in [n]$, Nagaev (1976) proves that

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \in A \right\} - \mathbb{P}\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Y}_i \in A \right\} \right| \geqslant c \frac{\mathbb{E}\left[\| Y_1 \|^3 \right]}{\sqrt{n}}, \tag{5.2}$$

where $\mathcal{A}$ is the class of all Borel measurable convex sets and $c > 0$ is some universal constant. This

last result, together with an simple application of Jensen's inequality

$$\mathbb{E}\left[\| Y_1 \|^3\right] \geqslant \left(\mathbb{E}\left[\| Y_1 \|^2\right]\right)^{3/2} = p^{3/2},$$

tells us that when $p \sim n^{1/3}$, the right-hand side of (5.2) is strictly bounded away from zero in the limiting regime $n \to \infty$, and hence a CLT simply cannot hold. Nevertheless, the groundbreaking work of Chernozhukov et al. (2013) shows that, instead of (5.2), one has that

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \max_{j\in[p]} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{ij} \leqslant t \right\} - \mathbb{P}\left\{ \max_{j\in[p]} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Y}_{ij} \leqslant t \right\} \right| \to 0 \quad \text{as } n \to \infty,$$

where p is allowed to be exponentially larger than n. This powerful result, which is very suitable for simultaneous hypothesis testing in high-dimensional regime, is then generalized and utilized by several papers to high-dimensional hypothesis testing problem under different weak dependence structures. Specifically, Zhang and Cheng (2014), Zhang and Wu (2017), and Zhang and Cheng (2018) focus on weakly dependent stationary time series that are characterized in terms of functional dependence, a measure of dependency introduced in Wu (2005); stationary β-mixing (absolute regular) sequence of random vectors is treated in the extension of Chernozhukov et al. (2019); while Belloni and Oliveira (2018) focuses on sequence of martingale differences.

Contribution and Organization of the Chapter

The main contribution of this chapter is to bridge the line of research in high-dimensional statistics and the theory of finite horizon MDPs, where dependency and non-stationarity are endogenously generated by decision makers. The rest of the chapter is organized as follows. We review related literature in Section 5.1. In Section 5.2, we introduce a general MDP framework that we will be working with. Then we take up the problems of hypothesis testing and the construction of simultaneous confidence intervals. Such problems are then treated in Section 5.3 and Section 5.4. Roughly speaking, Section 5.3 deals with policies that make states of MDPs "regenerative"; while Section 5.4 deals with MDPs with "transient" states. Then in Section 5.5, we consider a dynamic inventory control problem to which we apply the results of Section 5.3. Later in Section 5.6, we revisit the problem of sequential monotone subsequence selection from a random sample, that fits in the results of Section 5.4. Finally, we conclude in Section 5.7.

5.1 Literature Review

Markov decision problems, which also bear other names such as stochastic dynamic program, sequential stochastic optimization, and discrete time stochastic control, have been extensively studied in the literature. Many structural results have been obtained for the infinite-horizon formulation of such problems, in which the objective is usually the expected total reward with discount or the expected long-run average reward. In such formulations, the optimal policy is often stationary (time-independent). Asymptotic behaviors of the (random) total reward collected under stationary policies have been studied in great details in the literature. Substantial results have been attained through the works of Mandl and Laušmanová (1991); Mendoza-Pérez (2008); Mendoza-Pérez and Hernández-Lerma (2010). On the other hand, however, the finite-horizon case is more delicate and much less touched. A recent work that deserves attention is Arlotto and Steele (2016), who study a finite-horizon MDP, and establish the limiting distribution of the total reward collected under the optimal policy that is non-stationary (time-dependent).

The total reward collected under any sequential policy of a Markov decision problem can be viewed as the sum of the reward function on the states of the MDP. Since the states of the MDP form a Markov chain whose transition law is determined by the policy, the analysis of the total collected reward can also be analyzed through *Markov chain central limit theorems*. Maxwell and Woodroofe (2000) established asymptotic normality for partial sum of additive functionals on stationary ergodic Markov chain under moment conditions. Drift conditions are also proposed to guarantee central limit theorems on Markov chains, see, e.g. Jarner and Roberts (2002, Theorem 4.2) and Meyn and Tweedie (2009, Theorem 17.0.1). Again, a vast majority of this literature also focus on the stationary case. An excellent survey and extensive references on Markov chain CLT can be found in Jones (2004).

In addition to Markov chain central limit theorems, *mixing conditions* have also been widely used for establishing limiting results. The applications of such conditions include the context of Markov chains, as well as dependent sequences and time series. Most attention along this stream of literature has been put on the stationary case. Peligrad (1985, 1990) proves (functional) central limit theorems for stationary φ-mixing sequences under different moment conditions. Utev (1991) studies φ-mixing sequence and establishes a central limit theorem under a Lindeberg type condition. An exception that deserves special attention is Peligrad (2012), who proves a central limit theorem

for the sum of additive functionals on a non-homogeneous Markov chain whose ρ-mixing coefficient satisfies $\rho_1 < 1$. Another work for the non-stationary case is Fryzlewicz and Rao (2011), in which the mixing rates of time-varying ARCH time series models were obtained. An excellent review with more sustained references about different mixing conditions and their implications can be found in Bradley (2005).

All the above mentioned literature have been focused on proving limiting results in the classic setup where the dimension is fixed. However, as we discussed earlier, certain types of statistical inferences are still possible even when we are in the high-dimensional regime in which such above-mentioned results fail. This high-dimensional regime has drawn significant attention within the statistics literature. A series of remarkable results for the independent case have been achieved along the works of Chernozhukov et al. (2013, 2015, 2017). With these results, the applicability of Gaussian approximations have been pushed much further to the regime where the dimension of the estimand could be much larger than the sample size. See also the recent survey Belloni et al. (2018). Another notable recent work that deserves special attention is Belloni and Oliveira (2018), who develop high-dimensional central limit theorems and simultaneous inferences under martingale settings.

Different from the situation in low-dimension regime, the estimation of the covariance matrices under dependency is cumbersome when the dimension is large. As a result, the classic statistical inference approaches are not applicable any more. As a non-parametric procedure, *bootstrap* and its different variants become useful for high-dimensional inferences. In Chernozhukov et al. (2019), together with the block multiplier bootstrap method, β-mixing condition is used to construct simultaneous confidence intervals in high dimension. This particular bootstrap method can be thought of as a special version of the tapered bootstrap that has been studied in Paparoditis and Politis (2001, 2002) and Andrews (2004). We also refer interested readers to Lahiri (2003) as a general reference about resampling methods for dependent data. Different than β-mixing, Wu (2005) introduces another notion of dependence called functional dependence, and it is subsequently used in Zhang and Cheng (2014); Zhang and Wu (2017) and Zhang and Cheng (2018) to establish the validity of bootstrapping for stationary time series data in high dimension. We note that neither of the two weak-dependency conditions implies the other.

5.2 A General MDP Framework

We consider in total p decision makers, labeled by $j \in [p]$, each of whom faces a discrete-time finite horizon MDP with n periods, labeled by $t \in [n]$. Denote $\mathcal{X}_j$ as the *state space* for the j-th decision maker (referred to as *she*). During each period, she takes an action that will have impact on the future rewards she collects. To incorporate different applications, we assume that there are uncertainties realized both before and after each action is taken. Specifically, at the beginning of each time period $t \in [n]$, suppose her MDP is in state $x_{(t-1)j}$, there first realizes an exogenous uncertainty $y_{tj} \in \mathcal{Y}_j$, then she chooses an action $a_{tj} \in \mathcal{A}(x_{(t-1)j}, y_{tj})$. The policy π_j that she implements is represented through how such action is chosen. Here one can think of the policy π_j as the optimal policy or a near-optimal policy. After such action is taken, another uncertainty $w_{tj} \in \mathcal{W}_j$ is realized, then she obtains a reward $r_j(x_{(t-1)j}, y_{tj}, a_{tj}, w_{tj})$, and the state of her MDP involves as

$$x_{tj} = f_j(x_{(t-1)j}, y_{tj}, a_{tj}, w_{tj}),$$

where r_j and f_j are two deterministic functions. The total reward she collects under policy π_j is

$$R_{nj} = R_{nj}(\pi_j) = \sum_{t=1}^{n} r_j(X_{(t-1)j}, Y_{tj}, A_{tj}, W_{tj}). \tag{5.3}$$

Denote $v_{nj}(x)$ as the expected total reward she collects by implementing policy π_j when the initial state of her MDP is $X_{0j} = x$. That is,

$$v_{nj}(x) = \mathbb{E}\left[R_{nj}(\pi_j)\right].$$

Similarly, we define her expected reward-to-go $v_{(n-k)j}(x)$ as the expected total reward that she collects when there are $n-k$ periods remaining and the state of her MDP is $X_{kj} = x$. That is,

$$v_{(n-k)j}(x) = \mathbb{E}\left[\sum_{t=k+1}^{n} r_j(X_{(t-1)j}, Y_{tj}, A_{tj}^{\pi_j}, W_{tj}^{\pi_j}) \,\middle|\, X_{kj} = x\right], \tag{5.4}$$

where the notations A^{π_j} and W^{π_j} indicate that they are under policy π_j.

Sklar's Theorem and Copulas

We assume that all exogenous uncertainties faced by each decision maker have joint distribution F, and such uncertainties are independent across time periods. For any multi-dimensional distribution

F with marginal distributions $\{F_j\}_{j\in[p]}$, Sklar's theorem (Sklar, 1959) tells us that the joint distribution can be equivalently represented as a copula function of all the marginals. Specifically, one has that

$$F(w_1,\ldots,w_p) = C(F_1(w_1),\ldots,F_p(w_p)) \qquad \text{for all } (w_1,\ldots,w_p)\in\mathbb{R}^p.$$

The copula $C(\cdot):[0,1]^p\to[0,1]$ together with all the marginal distributions $\{F_j\}_{j\in[p]}$ totally determine the entire distribution F. However, an arbitrary unknown copula C could put complex unknown correlation structure on the total rewards collected by each decision maker, even if all the marginal distributions are known. In other words, assuming knowing marginal distributions of the uncertainty puts the problem into a non-trivial yet tractable context.

5.2.1 Hypothesis Testing and the Construction of Simultaneous Confidence Intervals

We are interested in testing the null hypothesis

$$H_0 = \{\text{ every decision maker } j\in[p] \text{ is implementing her designated policy } \pi_j\},$$

where the number of decision makers, p, is allowed to be comparable to or even much larger than the number of time periods, n. We are also interested in constructing simultaneous confidence intervals for the total rewards collected by each decision maker under the null hypothesis. That is, for the vector

$$\boldsymbol{R}_n \equiv (R_{n1}(\pi_1), R_{n2}(\pi_2),\ldots,R_{np}(\pi_p))'\in\mathbb{R}^p,$$

and prescribed confidence level α, we intend to construct simultaneous confidence intervals $\prod_{j=1}^p I_j(n,\alpha)$, with $I_j(n,\alpha)\subset\mathbb{R}$ for each $j\in[p]$, such that under the null hypothesis H_0,

$$\mathbb{P}\{R_{nj}\in I_j(n,\alpha) \text{ for all } j\in[p]\}\to 1-\alpha, \qquad \text{as } n\to\infty.$$

5.2.2 Martingale Representation

For the total reward collected by a policy of a finite horizon MDPs, one can define a Doob's martingale with respect to the filtration $\{\mathcal{F}_t\}_{t\in[n]}$, which in turn induces a sequence of martingale differences. Specifically, let $\mathcal{F}_0$ be the trivial σ-field and let $\mathcal{F}_t=\sigma\{Y_1,W_1,\ldots,Y_t,W_t\}$ for $t\in[n]$, then for all $t\in[n]$,

$$\boldsymbol{d}_t \equiv \mathbb{E}[\boldsymbol{R}_n \mid \mathcal{F}_t] - \mathbb{E}[\boldsymbol{R}_n \mid \mathcal{F}_{t-1}]\in\mathbb{R}^p$$

compose a sequence of martingale differences in $\mathbb{R}^p$ that satisfy that $\mathbb{E}[\boldsymbol{d}_t \mid \mathcal{F}_{t-1}] = 0$ for all $t \in [n]$ and

$$\sum_{t=1}^{n} \boldsymbol{d}_t = \boldsymbol{R}_n - \mathbb{E}[\boldsymbol{R}_n]. \tag{5.5}$$

Since each martingale difference $\boldsymbol{d}_t$ is $\mathcal{F}_t$-measurable, from the total reward representation (5.3) and the value function representation (5.4), we have another useful representation that for the j-th coordinate of $\boldsymbol{d}_t$,

$$d_{tj} = v_{(n-t)j}(X_{tj}) + r_j(X_{tj}, Y_{tj}, A_{tj}, W_{tj}) - v_{(n-t+1)j}(X_{(t-1)j}) \quad \text{for all } j \in [p]. \tag{5.6}$$

We close this section with the observation that with this last representation, we can construct test statistics for our statistical inference problems. Denote $\widehat{d}_{tj}$ as the realization of the martingale difference d_{tj}. That is, if we let the tuple $(x_{(t-1)j}, y_{tj}, a_{tj}, w_{tj})$ be the *realization* of $(X_{(t-1)j}, Y_{tj}, A_{tj}, W_{tj})$, then the state x_{tj} is determined as

$$x_{tj} = f_j(x_{(t-1)j}, y_{tj}, a_{tj}, w_{tj}),$$

and so we have the realized (observed) martingale difference

$$\widehat{d}_{tj} = v_{(n-t)j}(x_{tj}) + r_j(x_{(t-1)j}, y_{tj}, a_{tj}, w_{tj}) - v_{(n-t+1)j}(x_{(t-1)j}). \tag{5.7}$$

The sequence of the realized martingale differences will serve as the building block of our bootstrap procedure later.

5.3 MDPs with Regenerative Property

In certain types of MDPs, the system states under certain policies enjoy regenerations in the sense that with high probability, the system state becomes independent of past history that is far away. In other words, the dependency between the states at two distant time periods decays as the distance gets large. When the speed of such decay is well balanced with the dimension p, we are able to conduct a block version of multiplier bootstrap method to obtain critical value for simultaneous hypothesis testing and to construct simultaneous confidence intervals. The decaying condition we need is made precise in the following definition.

Definition 5.1 (Tail of regeneration time.). Let $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\} \subset \prod_{j\in[p]} \mathcal{X}_j$ be the sequence of (joint) states of each decision maker, that is,

$$\boldsymbol{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tp}) \in \prod_{j\in[p]} \mathcal{X}_j \quad \text{for all } t = 0, 1, \ldots, n.$$

We say that the MDPs satisfy regenerative property with rate $\varphi(p, k)$ if for each t, there exists a random regeneration time τ such that $\boldsymbol{X}_{\tau+t}$ is independent of $\boldsymbol{X}_t$, and the tail probability of τ satisfies $\mathbb{P}\{\tau > k\} \leqslant \varphi(p, k)$.

The interplay between p and k in the decay rate function $\varphi(p, k)$ is the driving force of how large p is allowed for our statistical inferences. We close this part with the following remark before diving into the multiplier bootstrap procedure for testing our null hypothesis.

Remark 5.1. (*φ-mixing vs. β-mixing.*) If the function $\varphi(p, k) \to 0$ as $k \to \infty$, then the sequence $\{\boldsymbol{X}_t\}_{t\in[n]}$ is φ-mixing (see, e.g., Fan and Yao, 2003, Section 2.6). In fact, this mixing condition can be further relaxed to β-mixing. However, β-mixing condition is usually difficult to verify in applications, while on the other hand, the regeneration condition is common and much easier to verify in certain concrete MDPs.

5.3.1 Test Statistic and the Block Multiplier Bootstrap

Often times, for MDP policies that induce regenerative states, the variance of the total reward they collect grows linearly in terms of time span. To cope this linear growth of the variance, we use the regular $n^{-1/2}$-normalization and the realized martingale differences from (5.7) to construct the test statistic

$$\widehat{T} \equiv \max_{j\in[p]} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{d}_{tj}. \tag{5.8}$$

Unfortunately, the correlation structure of the martingale differences across decision makers is unknown, and hence the computation of the critical value is not straightforward. In order to overcome this obstacle, we use block multiplier bootstrap method, with the big-block-small-block idea that dates back to Bernstein (1927). We note that this block scheme has been used in Yu (1994); Chen et al. (2016); Chernozhukov et al. (2019) to prove various convergence results under different weak dependencies.

The big-block-small-block idea is simple. It splits the entire sequence of the martingale differences into big blocks and small blocks alternatively, and drops the small blocks away. When the inter-dependency of the original sequence decays fast enough in terms of the distance between blocks, the resulted big blocks can be well approximated by a corresponding sequence of independent random variables. Mathematically, we first choose two block sizes $m_1(n)$ and $m_2(n)$ for big blocks and small blocks respectively, such that as $n \to \infty$, they satisfy that $m_1(n) = o(n)$ and $m_2(n) = o(m_1(n))$ while $m_2(n) \to \infty$. Then we set $q = \lfloor n/(m_1 + m_2) \rfloor$ and divide the horizon $\{1, \ldots, n\}$ into alternating groups of size m_1 and m_2. That is, for all $k \in [q]$, we have

$$I_k = \{(k-1)(m_1 + m_2) + 1, \ldots, (k-1)(m_1 + m_2) + m_1\},$$
$$J_k = \{(k-1)(m_1 + m_2) + m_1, \ldots, k(m_1 + m_2)\},$$

and $J_{q+1} = \{q(m_1 + m_2) + 1, \ldots, n\}$. Recalling the realized martingale differences from (5.7), we define shorthand $\widehat{\mu}_j$ for their sample average over big blocks as

$$\widehat{\mu}_j \equiv \sum_{k=1}^{q} \sum_{t \in I_k} \widehat{d}_{tj}.$$

Next, we generate a sequence of i.i.d. standard normal random variables $\{\epsilon_k\}_{k \in [q]}$ that are independent of everything else, and use them together with (5.7) to construct multiplier bootstrap statistic

$$\widehat{W} \equiv \max_{j \in [p]} \left| \frac{1}{\sqrt{m_1 q}} \sum_{k=1}^{q} \epsilon_k \sum_{t \in I_k} \left(\widehat{d}_{tj} - \widehat{\mu}_j \right) \right|.$$

Then for any given level of significance α, we use the critical value

$$\widehat{c}(\alpha) \equiv (1-\alpha)\text{-quantile of } \widehat{W} \text{ given } \widehat{\boldsymbol{d}}_1. \tag{5.9}$$

That is, we reject the null hypothesis H_0 if our test statistic from (5.8) $\widehat{T} > \widehat{c}(\alpha)$. Based on the critical value $\widehat{c}(\alpha)$, we also have the simultaneous confidence intervals for the total rewards $\boldsymbol{R}_n \in \mathbb{R}^p$ under H_0 given as

$$I_j(n, \alpha) = \left[v_{nj}(x_{0j}) + \sum_{t=1}^{n} \widehat{d}_{tj} - \widehat{c}(\alpha)\sqrt{n}, v_{nj}(x_{0j}) + \sum_{t=1}^{n} \widehat{d}_{tj} + \widehat{c}(\alpha)\sqrt{n} \right] \quad \text{for all } j \in [p].$$

5.3.2 Asymptotic Validity

The asymptotic validity of the block multiplier bootstrap procedure is given in the following theorem, which is adopted from Chernozhukov et al. (2019, Theorem B.1).

Theorem 5.1 (Testing for MDPs with regenerative states). *Consider p (possibly) correlated MDPs described as in Section 5.2. Suppose that there exist constants c_0, c_1, c_2, C_1 such that the following conditions hold.*

(i) Uniform bound on the martingale differences.

$$\| \boldsymbol{d}_t \|_\infty \leqslant C_1 < \infty \quad \text{and} \quad \min_{j \in [p]} \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1} \right] \geqslant c_1, \quad \text{for all } t \in [n].$$

(ii) Tail of regeneration time. The MDPs satisfy the regenerative property defined in Definition 5.1, with tail of the regeneration time satisfying

$$\mathbb{P}\{\tau > k\} \leqslant \varphi(p, k) \leqslant p \exp(-c_0 k).$$

(iii) The parameters satisfy:

$$\max\left\{ \frac{np}{m_1 + m_2} \exp(-c_0 m_2), \frac{m_2}{m_1} \log^2 p \right\} \leqslant C_1 n^{-c_2}, \quad \text{and} \quad m_1 \log^{5/2}(pn) \leqslant C_1 n^{1/2 - c_2}. \tag{5.10}$$

Then under the null hypothesis H_0 in (5.1)*, we have that the critical value* (5.9) *generated by the block multiplier bootstrap satisfies that*

$$\left| \mathbb{P}\left\{ \widehat{T} > \widehat{c}(\alpha) \right\} - \alpha \right| \leqslant C n^{-c},$$

with the two constants c, C depending only on c_0, c_1, c_2, C_1.

5.4 MDPs with Absorbing States

The other special class of MDPs is that with *absorbing state*. Such absorbing state usually makes the variance of the total reward collected by the optimal policy and near-optimal policies grow sublinearly in the length of time horizon. Specifically, recall the martingale differences given in (5.5), that is,

$$\sum_{t=1}^{n} \boldsymbol{d}_t = \boldsymbol{R}_n - \mathbb{E}\left[\boldsymbol{R}_n \right].$$

In this section we focus on MDPs in which the variance of R_{nj} grows sublinearly in n for all $j \in [p]$. The subtle differences between this class of MDPs and the class of MDPs with regenerative property make our previous analysis inapplicable. Fortunately, however, when more knowledge of the martingale structure is available, one can still conduct bootstrap procedure to obtain critical value for simultaneous hypothesis testings and to construct simultaneous confidence intervals for the total rewards collected by each decision maker that is implementing her designated policy.

5.4.1 Test Statistics and the Multiplier Bootstrap

We start with constructing test statistics with *self-normalization*. Let $\sigma_{nj}^2 = n^{-1}\text{Var}\left\{\sum_{t=1}^n d_{tj}\right\}$ be the normalized variance. Then we normalize the martingale differences by square root of this variance to obtain

$$Z_{tj} \equiv \frac{d_{tj}}{\sigma_{nj}}, \quad \text{for all } t \in [n] \text{ and } j \in [p],$$

so that the associated variance is

$$\text{Var}\left\{\sum_{t=1}^n Z_{tj}\right\} = \frac{1}{\sigma_{nj}^2}\text{Var}\left\{\sum_{t=1}^n d_{tj}\right\} = n, \text{ for all } j \in [p].$$

Then from the realized martingale differences (5.7), we obtain the sample variance

$$\widehat{\sigma}_{nj}^2 \equiv \frac{1}{n}\sum_{t=1}^n \widehat{d}_{tj}^2 \quad \text{for all } j \in [p].$$

With this last sample variance, we can further obtain the realized self-normalized quantity

$$\widehat{Z}_{tj} \equiv \frac{\widehat{d}_{tj}}{\widehat{\sigma}_{nj}}, \quad \text{for all } t \in [n] \text{ and } j \in [p].$$

In principle, with the knowledge of all marginal distributions of the exogenous uncertainty of the possibly correlated MDPs, it is possible to know the exact marginal distributions of the vector of total rewards $\boldsymbol{R}_n \in \mathbb{R}^p$. However, the covariance structure of $\boldsymbol{R}_n$ is unknown since the correlation structure of the exogenous uncertainty is unknown. To overcome this, we use the following bootstrap procedure to generate the critical value. We generate i.i.d. standard normal random variables $\{\epsilon_t\}_{t\in[n]}$ that are independent from everything else, and let

$$\widehat{Z} = \max_{j\in[p]} \left| \frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t \widehat{Z}_{tj} \right|.$$

Then we set the critical value as

$$\widehat{c}(\alpha) \equiv (1-\alpha)\text{-quantile of } \widehat{Z} \text{ conditional on } \{\widehat{Z}_{1j}\}_{j\in[p]}.$$

Based on this test statistics, we reject the null hypothesis H_0 if out test statistics

$$\widehat{T} \equiv \max_{j\in[p]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{Z}_{tj} \right| > \widehat{c}(\alpha).$$

Moreover, we have the simultaneous confidence intervals for the vector of total rewards $\boldsymbol{R}_n \in \mathbb{R}^p$, given as

$$\left[v_{nj}(x_{0j}) + \sum_{t=1}^{n} \widehat{d}_{tj} - \widehat{c}(\alpha)\sqrt{n}\widehat{\sigma}_{nj}, v_{nj}(x_{0j}) + \sum_{t=1}^{n} \widehat{d}_{tj} + \widehat{c}(\alpha)\sqrt{n}\widehat{\sigma}_{nj} \right] \quad \text{for all } j \in [p]. \tag{5.11}$$

5.4.2 Asymptotic Validity

The above bootstrap procedure for hypothesis testing and construction of simultaneous confidence intervals is asymptotically valid under suitable conditions. The asymptotic validity is mainly based on Belloni and Oliveira (2018, Assumption 2), and we tailor the conditions to adapt to our setup and notations. As we will see in Section 5.6, it is usually not trivial to verify these conditions in applications, and such verifications often require a substantial understanding of the MDPs at hand. The asymptotic validity result is made precise in the following statement.

Theorem 5.2 (Testing for MDPs with absorbing states). *Let δ_n and ψ_n be two fixed sequences that are going to zero as n increases. Suppose that the martingale differences defined in* (5.6) *satisfy the following conditions.*

(i) $\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\max_{j\in[p]} |\, Z_{tj} \,|^3\right] \leqslant n^{\rho}$ *for some* $\rho \in (0, 1/2)$*;*

(ii) The matrices $V \equiv \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\boldsymbol{Z}_t \boldsymbol{Z}_t'\right]$ *and* $V_n \equiv \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\boldsymbol{Z}_t \boldsymbol{Z}_t' \mid \mathcal{F}_{t-1}\right]$ *satisfy that*

$$\max_{i,j\in[p]} |\, V_{ij} \,| \leqslant C \quad \text{and} \quad \mathbb{E}\left[\max_{i,j\in[p]} |\, V_{n,ij} - V_{ij} \,|\right] \leqslant \frac{\delta_n}{\log^2(np)},$$

and with probability at least $1-\psi_n$*, we have that*

$$V_n \preceq \left(1 + \delta_n/\log^2 p\right) V, \quad \text{and} \quad \max_{i,j\in[p]} |\, V_{n,ij} - V_{ij} \,| \leqslant \frac{\delta_n}{\log^2(np)};$$

(iii) With probability at least $1-\psi_n$*, we have that*

$$\max_{i,j\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n} Z_{ti}Z_{tj} - \mathbb{E}\left[Z_{ti}Z_{tj} \mid \mathcal{F}_{t-1}\right]\right| \vee \max_{j\in[p]}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{Z}_{tj} - Z_{tj}\right)^2\right| \leqslant \frac{\delta_n^2}{\log^2(np)};$$

(iv) $\psi_n \leqslant \delta_n^{1/2}$*,* $\log^{7/2} p \leqslant \delta_n n^{1/2-\rho}$*, and* $\min_{j\in[p]}\sigma_j \geqslant c$*.*

Then the simultaneous confidence intervals in (5.11) *are asymptotic valid. That is, as* $n\to\infty$*,*

$$\left|\mathbb{P}\left\{\sum_{t=1}^{n}\widehat{d}_{tj} - \widehat{c}(\alpha)\sqrt{n}\widehat{\sigma}_{nj} \leqslant R_{nj} - v_{nj}(x_{0j}) \leqslant \sum_{t=1}^{n}\widehat{d}_{tj} + \widehat{c}(\alpha)\sqrt{n}\widehat{\sigma}_{nj},\ \textit{for all } j\in[p]\right\} - (1-\alpha)\right| \to 0.$$

5.5 Application: Testing Stationary Inventory Control Policies

To illustrate how our MDP framework and the multiplier bootstrap procedure in Section 5.3 can be applied to concrete applications, we consider *dynamic inventory control problems*. The setup we will be focusing on is in which the decision makers are implementing the stationary policies that are optimal to the infinite horizon problem. While we observe a time horizon of finite length, and we are interested in simultaneously testing whether each of the decision makers is implementing her designated policy.

5.5.1 Problem Setup (Scarf, 1959; Iglehart, 1963).

We now describe a dynamic inventory control problem faced in parallel by multiple decision makers. Suppose there are p decision makers each of whom manages a inventory of a single product over an infinite time horizon, and we get to observe the costs they incur and their each-day inventory levels over a time horizon of the length n. For each decision maker $j\in[p]$, the inventory level starts from initial level X_{0j}, and she implements a stationary policy that does not change over time. Then in each time period $t\in[n]$, she needs to decide whether to order the product or not, and how many units of the product to order if she decides to place an order. If we denote $\gamma_j(x)$ as her order-up-to function, that is, she will order $(\gamma_j(x)-x)$ units of the product at time period t if her inventory level from the last period is $X_{(t-1)j}=x$. Then it is assumed that the order will be fulfilled right away and an uncertain demand ξ_{tj} will be realized after the order is delivered (if she places an order), so

that the inventory level evolves as

$$X_{tj} = \gamma_j(X_{(t-1)j}) - \xi_{tj}. \tag{5.12}$$

The marginal distributions of ξ_{tj}'s for all $j \in [p]$ are known. However, the correlation among different coordinates of the vector $\boldsymbol{\xi}_t$ of demands is unknown and could be arbitrary. There are three types of costs. The ordering cost

$$c_{o,j}(x) = \begin{cases} K + c_j x & \text{if } x > 0 \\ 0 & \text{if } x = 0, \end{cases} \tag{5.13}$$

the per-unit holding cost $c_{h,j}$, and the per-unit backlog penalty $c_{p,j}$. More specifically, there is a fixed ordering cost K, and the j-th individual decision maker needs to pay $K + c_j(y - x)$ to bring the inventory of her product from x up to y before the demand realization. Then after the demand is realized, if the remaining inventory level x is positive, then she incurs a holding cost $c_{h,j}x$ and these x units of product will be carried over to the next period; while if the inventory level x is negative, she incurs a backlog penalty $-c_{p,j}x$ for not being able to meet these demands. The latter two types of costs can be written in one single function as

$$L_j(x) = \begin{cases} c_{h,j}x & \text{if } x \geqslant 0 \\ -c_{p,j}x & \text{if } x < 0. \end{cases} \tag{5.14}$$

All the decision makers are implementing the optimal stationary base-stock policy for the infinite horizon problem (Iglehart, 1963). That is, for each decision maker $j \in [p]$, there is a pair of base-stock levels (s_j, S_j) with $s_j < S_j$ such that whenever the inventory falls below s_j, she will place an order and bring the inventory level up to S_j. Equivalently, she has the order-up-to function given by

$$\gamma_j(x) = \begin{cases} S_j & \text{if } x \leqslant s_j \\ x & \text{if } x > s_j. \end{cases} \tag{5.15}$$

The total cost she incurs by implementing this base-stock policy over a time horizon of length n is given as

$$R_{nj} = \sum_{t=1}^{n} \left\{ c_{o,j}(\gamma_j(X_{(t-1)j}) - X_{(t-1)j}) + L_j(X_{tj}) \right\}.$$

We also have the value function representation

$$v_{nj}(X_{0j}) = \mathbb{E}\left[R_{nj}\right],$$

and for all $t \in [n]$,

$$v_{(n-t)j}(X_{t,j}) = \mathbb{E}\left[\sum_{\ell=t}^{n} \left\{c_{o,j}(\gamma_j(X_{(\ell-1)j}) - X_{(\ell-1)j}) + L_j(X_{\ell j})\right\}\right].$$

As a result, our martingale differences in this context are

$$\begin{aligned} d_{tj} =& \mathbb{E}\left[R_{nj} \mid \mathcal{F}_t\right] - \mathbb{E}\left[R_{nj} \mid \mathcal{F}_{t-1}\right] \\ =& c_{o,j}(\gamma_j(X_{(t-1)j}) - X_{(t-1)j}) + L(X_{tj}) + v_{(n-t)j}(X_{tj}) - v_{(n-t+1)j}(X_{(t-1)j}). \end{aligned}$$

With the shorthand $\widehat{v}_k(x) = L(x) + v_k(x)$, one can easily obtain that

$$d_{tj} = \widehat{v}_{(n-t)j}(\gamma_j(X_{(t-1)j}) - \xi_{tj}) - \mathbb{E}\left[\widehat{v}_{(n-t)j}(\gamma_j(X_{(t-1)j}) - \xi_{tj}) \,\middle|\, \mathcal{F}_{t-1}\right]. \tag{5.16}$$

The marginal distributions of demands play a crucial role in the analysis, and we put the certain regularity conditions on these marginal distributions. For all the base-stock level pairs (s_j, S_j), $j \in [p]$, let $\underline{s} = \min_{j\in[p]} s_j$ and $\bar{s} = \max_{j\in[p]} S_j - s_j$. The regularity conditions we need are the following.

Definition 5.2 (Regular class of distributions). A p-dimensional non-negative bounded random variable $\boldsymbol{W} \in \mathbb{R}^p$ with continuous distribution $F(\cdot) : [0, \bar{D}]^p \to [0, 1]$ and marginal densities f_j for all $j \in [p]$, is said to be *regular* if the following conditions are satisfied.

- Finite, non-vanishing means:

$$\underline{\mu} \equiv \min_{j\in[p]} \mathbb{E}\left[W_j\right], \quad \bar{\mu} \equiv \max_{j\in[p]} \mathbb{E}\left[W_j\right], \quad \text{and} \quad 0 < \underline{\mu} \leqslant \bar{\mu} < \infty.$$

- Finite variance:

$$\max_{j\in[p]} \operatorname{Var}\left\{W_j\right\} \leqslant \bar{\sigma}^2 < \infty.$$

- Non-flat marginal density: for all the marginal distribution F_j with density f_j,

$$\max_{j\in[p]} \left\{\mathbb{P}\left\{W_j \leqslant S_j\right\}, \mathbb{P}\left\{W_j \geqslant s_j\right\}\right\} \leqslant \bar{p} < 1 \quad \text{and} \min_{j\in[p]} \inf_{w\in[0,\bar{D}]} f_j(w) \geqslant \eta > 0. \tag{5.17}$$

- For each marginal distribution with density f_j, and for each $\epsilon \geqslant 0$, there is $\widehat{w} = \widehat{w}(\epsilon)$ such that

$$\begin{aligned} f_j(w) - f_j(w+\epsilon) \leqslant 0 \quad & \text{for all } w \leqslant \widehat{w}, \text{ and} \\ f_j(w) - f_j(w+\epsilon) \geqslant 0 \quad & \text{for all } w \geqslant \widehat{w}. \end{aligned}$$

The boundedness of the means and variances essentially rules out trivial cases in which demands are deterministic. The last condition is borrowed from Arlotto and Steele (2016, Definition 12), which is used to guarantee boundedness of the contraction coefficient there. We will use similar analysis on the individual inventory levels.

If the joint uncertain demands faced by each of the decision makers satisfy the regularity conditions given in Definition 5.2, then we can apply the bootstrap procedure from Section 5.3 to construct simultaneous confidence intervals for the total costs incurred by each decision maker with asymptotic validity. Such results are summarized in the following theorem.

Theorem 5.3. *If the joint demand distribution F is regular as defined in Definition 5.2, then the simultaneous confidence intervals*

$$\prod_{j=1}^{p}\left[v_{nj}(x_{0j})+\sum_{t=1}^{n}\widehat{d}_{tj}-\widehat{c}(\alpha)\sqrt{n},v_{nj}(x_{0j})+\sum_{t=1}^{n}\widehat{d}_{tj}+\widehat{c}(\alpha)\sqrt{n}\right]$$

as constructed in Section 5.3 are asymptotic valid. That is,

$$\mathbb{P}\left\{v_{nj}(x_{0j})+\sum_{t=1}^{n}\widehat{d}_{tj}-\widehat{c}(\alpha)\sqrt{n}\leqslant R_{nj}\leqslant v_{nj}(x_{0j})+\sum_{t=1}^{n}\widehat{d}_{tj}+\widehat{c}(\alpha)\sqrt{n}\right\}\to(1-\alpha),\qquad \text{as } n\to\infty.$$

5.5.2 Regeneration Time with Exponential Tail

We consider the regeneration time $(\tau+1)$ which is the first time that every decision maker has at least put an order. We establish an exponentially decaying bound for the tail probability of τ, which is a stopping time of the Markov chain $\{\boldsymbol{X}_0,\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\}\subset\mathbb{R}^p$ of the aggregated inventory levels induced by each individual decision maker's policy. The basic idea is that each time a decision maker places an order, her inventory level regenerates in the sense that the future evolution of her inventory level becomes independent of the past history. If one looks at the particular stopping time τ, which is the first time that for every decision maker j, her inventory level has fall below s_j for at least once, then some moments conditions on the marginal distributions of the uncertainty imply an exponentially decaying right tail probability of τ.

To establish this exponential decaying tail, we need the following lemma which concerns lower tail probability of sum of independent non-negative random variables. It can be obtained from several different well-known concentration inequalities. We provide a proof for the sake of completeness.

Lemma 5.1 (Lower tail of independent non-negative random variables). *Let $X_1, \ldots, X_n$ be independent non-negative random variables. Then for any $t > 0$,*

$$\mathbb{P}\left\{\sum_{i=1}^{n}\left(X_i - \mathbb{E}\left[X_i\right]\right) \leqslant -t\right\} \leqslant \exp\left(\frac{-t^2}{2\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]}\right).$$

Proof. For any non-negative random variable X, the simple inequality $e^{-u} \leqslant 1 - u + u^2/2$ for all $u \geqslant 0$ tells us that for all $\lambda > 0$,

$$\mathbb{E}\left[e^{-\lambda X}\right] \leqslant 1 - \lambda\mathbb{E}\left[X\right] + \lambda^2\mathbb{E}\left[X^2\right]/2.$$

Another elementary inequality $\log u \leqslant u - 1$ for all $u > 0$ gives us that

$$\log\mathbb{E}\left[e^{-\lambda X}\right] \leqslant \mathbb{E}\left[e^{-\lambda X}\right] - 1,$$

which in turn implies that

$$\log\mathbb{E}\left[e^{-\lambda X}\right] \leqslant -\lambda\mathbb{E}\left[X\right] + \lambda^2\mathbb{E}\left[X^2\right]/2.$$

If we take exponential of both sides and rearrange, we obtain that for all $\lambda > 0$,

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leqslant e^{\lambda^2\mathbb{E}\left[X^2\right]/2}.$$

This last inequality tells us that when we have independent non-negative random variables $X_1, \ldots, X_n$, the independence implies that

$$\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}\right] = \prod_{i=1}^{n}\mathbb{E}\left[e^{\lambda(X_i - \mathbb{E}[X_i])}\right] \leqslant e^{\lambda^2\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]/2}.$$

By Markov inequality, we know that for all $t > 0$,

$$\mathbb{P}\left\{\sum_{i=1}^{n}\left(X_i - \mathbb{E}\left[X_i\right]\right) \leqslant -t\right\} \leqslant e^{-\lambda t}\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}\right] \leqslant \exp\left\{-\lambda t + \lambda^2\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]/2\right\}.$$

Note that the last right-hand side is minimized by choosing $\lambda = t/\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]$, which in turn gives us the final tail probability bound

$$\mathbb{P}\left\{\sum_{i=1}^{n}\left(X_i - \mathbb{E}\left[X_i\right]\right) \leqslant -t\right\} \leqslant \exp\left(\frac{-t^2}{2\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]}\right),$$

just as we needed. □

With this last lemma, we are able to obtain an exponential-decay upper bound on the right tail probability of the stopping time τ. In fact, it is without loss of generality to assume the initial inventory $X_{0j} = S_j$ for all $j \in [p]$. General values of X_{0j} can be analyzed similarly. The tail bound is given in the following proposition.

Proposition 5.1 (Tail bound of the mixing time of joint inventory levels). *There exist constants K_0 and $\bar{s}$ that depend only on the marginal distributions of $F(\cdot) : \mathbb{R}_+^p \to [0,1]$ such that for $k \geqslant K_0$,*

$$\mathbb{P}\{\tau > k\} \leqslant p \exp\left(-\frac{\left(k\underline{\mu} - \bar{s}\right)^2}{2k\left(\bar{\mu}^2 + \bar{\sigma}^2\right)}\right),$$

with the constants $\underline{\mu}, \bar{\mu}, \bar{\sigma}^2$ given in Definition 5.2.

Proof. Recall that the structure of the (s, S) base-stock policy, and that $\bar{s} \equiv \max_{j\in[p]}(S_j - s_j)$, we have the inclusion

$$\{\omega : \tau > k\} \subset \bigcup_{j\in[p]} \left\{\omega : \sum_{t=1}^{k} \xi_{tj} < \bar{s}\right\}. \tag{5.18}$$

Therefore, we have the tail probability bound

$$\mathbb{P}\{\tau > k\} \leqslant \mathbb{P}\left\{\bigcup_{j\in[p]} \left\{\sum_{t=1}^{k} \xi_{tj} < \bar{s}\right\}\right\} \leqslant \sum_{j\in[p]} \mathbb{P}\left\{\sum_{t=1}^{k} \xi_{tj} < \bar{s}\right\}.$$

For each summand, we apply Lemma 5.1 to obtain that

$$\mathbb{P}\left\{\sum_{t=1}^{k} \xi_{tj} < \bar{s}\right\} = \mathbb{P}\left\{\sum_{t=1}^{k} (\xi_{tj} - \mu_j) < -(k\mu_j - \bar{s})\right\} \leqslant \exp\left(-\frac{(k\mu_j - \bar{s})^2}{2k(\mu_j^2 + \sigma_j^2)}\right),$$

where in the last step a lower-bound on $k \geqslant K_0$ with $K_0 \equiv \bar{s}/\underline{\mu}$ is needed ($\underline{\mu}$ defined in Definition 5.2) to guarantee the positiveness $k\mu_j - \bar{s} > 0$. In other words, this tail probability bound is valid for $k \geqslant K_0$. Finally, some simple algebra gives us the inequality

$$-\frac{(k\mu_j - \bar{s})^2}{2k(\mu_j^2 + \sigma_j^2)} \leqslant -\frac{\left(k\underline{\mu} - \bar{s}\right)^2}{2k\left(\bar{\mu}^2 + \bar{\sigma}^2\right)},$$

which completes the proof. □

Now we are left with proving that upon regeneration, $\boldsymbol{X}_\tau$ becomes independent of the initial state $\boldsymbol{X}_0$. Note that under the base-stock policy, the states $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\} \subset \mathbb{R}^p$ of the inventory levels form a Markov chain that evolves as that for all $t \in [n]$ and all $j \in [p]$,

$$X_{tj} = \begin{cases} X_{(t-1)j} - \xi_{tj} & \text{if } X_{(t-1)j} > s_j \\ S_j - \xi_{tj} & \text{if } X_{(t-1)j} \leqslant s_j. \end{cases} \tag{5.19}$$

In addition to the stopping time τ, the base-stock policy also induces, for each decision maker $j \in [p]$, a sequence of stopping times $0 \equiv \kappa_0^j < \kappa_1^j < \ldots < \kappa_k^j \leqslant n$, at which orders are placed. Specifically, by setting $\kappa_0^j = 0$, one can recursively define the stopping times

$$\kappa_k^j = \max \left\{ t : \kappa_{k-1}^j < t \leqslant n \text{ and } \sum_{\ell=\kappa_{k-1}^j}^{t} \xi_{\ell j} \leqslant S_j - s_j \right\} \quad \text{for } k \geqslant 1. \tag{5.20}$$

In addition to these stopping times, we also have the equivalent representation of τ as

$$\tau = \min \left\{ t \in [n] \cup \{\infty\} : \kappa_1^j \geqslant t \text{ for all } j \in [p] \right\}. \tag{5.21}$$

Then for each $m \in [n]$, we set $\boldsymbol{\eta}(m) \in \mathbb{R}^p$ as

$$\eta_j(m) = \max \left\{ \kappa_k^j \leqslant m : k \geqslant 0 \right\} \quad \text{for all } j \in [p]. \tag{5.22}$$

In words, the j-th element of the vector $\boldsymbol{\eta}(m)$ is the last time before m that the inventory level of the j-th decision maker falls below s_j. For a fixed integer $m \geqslant 1$, we define the configuration set $\Gamma(m)$ as

$$\Gamma(m) \equiv \left\{ \boldsymbol{t} \in [m]^p : \max_{j \in [p]} t_j = m \right\}.$$

That is, the elements of $\Gamma(m)$ all have the form $\boldsymbol{t} = (t_j)_{j \in [p]}$ where $t_j \in [m]$ for all $j \in [p]$ and the largest coordinate of $\boldsymbol{t}$ is equal to m. We call an element $\boldsymbol{t} \in \Gamma(m)$ a configuration. The structure of the base-stock policy implies that, with the knowledge of $\{\tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t}\}$, the state $\boldsymbol{X}_m$ becomes independent of the initial state $\boldsymbol{X}_0$. This property is summarized in the following lemma.

Lemma 5.2. *For all $m \in [n-1]$, all configuration $\boldsymbol{t} \in \Gamma(m)$, and all $\boldsymbol{x}_0 \in \mathbb{R}^p$, if $\boldsymbol{\eta}(m) \in [m]^p$ is the vector of the last time each decision maker puts an order, then*

$$\mathbb{P}\left\{ \boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \mid \boldsymbol{X}_0 = \boldsymbol{x}_0 \right\} = \mathbb{P}\left\{ \boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \right\}, \quad \textit{for all } B \subset \mathbb{R}^p.$$

Proof. The structure of the base-stock policy (5.19) tells us that, for each $j \in [p]$, whenever there is an order placed at some stopping time κ as defined in (5.20) (that is, $X_{\kappa j} \leqslant s_j$), the next-period inventory level depends only on the base-stock level S_j and the demand $\xi_{(\kappa+1)j}$. Moreover, the state X_{tj} before the next backorder is independent of X_{0j} in the sense that for all $k \geqslant 1$, all $\kappa_{k-1}^j < t \leqslant \kappa_k^j$, and all $B \subset (s_j, \infty)$,

$$\mathbb{P}\left\{X_{tj} \in B, \kappa_{k-1}^j = \kappa \;\middle|\; X_{0j}\right\} = \mathbb{P}\left\{\sum_{\ell=\kappa+1}^{t} \xi_{\ell j} \in S_j - B\right\} = \mathbb{P}\left\{X_{tj} \in B, \kappa_{k-1}^j = \kappa\right\}.$$

Similarly, for all $m \in [n-1]$, and any $\boldsymbol{t} \in \Gamma(m)$, we have that for all $B \subset \mathbb{R}^p$,

$$\begin{aligned}\mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \mid \boldsymbol{X}_0\right\} &= \mathbb{P}\left\{\sum_{\ell=t_j}^{m} \xi_{\ell j} \in S_j - B_j, j \in [p]\right\} \\ =&\mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t}\right\},\end{aligned}$$

so the proof is complete. □

With this last lemma, we are able to prove the regenerative property of $\boldsymbol{X}_t$ at the stopping time τ in (5.21).

Lemma 5.3 (Independence upon regeneration). *The inventory level after the stopping time τ is independent of the initial level $\boldsymbol{X}_0$ in the sense that*

$$\mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B \mid \boldsymbol{X}_0\right\} = \mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B\right\} \quad \forall B \subset \mathbb{R}^p.$$

Proof. We start with the conditional probability $\mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\}$ and note that by conditioning on τ and $\boldsymbol{\eta}(\tau)$, we have that for all $\boldsymbol{x}_0 \in \mathbb{R}^p$,

$$\begin{aligned}\mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\} &= \sum_{m=1}^{\infty} \sum_{\boldsymbol{t} \in \Gamma(m)} \mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\} \\ &= \sum_{m=1}^{\infty} \sum_{\boldsymbol{t} \in \Gamma(m)} \mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\}.\end{aligned}$$

From Lemma 5.2 we know that each summand in this last expression satisfies that

$$\mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t} \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\} = \mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t}\right\}.$$

Combining the last two identities, we obtain that

$$\mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B \mid \boldsymbol{X}_0 = \boldsymbol{x}_0\right\} = \sum_{m=1}^{\infty} \sum_{\boldsymbol{t} \in \Gamma(m)} \mathbb{P}\left\{\boldsymbol{X}_{m+1} \in B, \tau = m, \boldsymbol{\eta}(m) = \boldsymbol{t}\right\} = \mathbb{P}\left\{\boldsymbol{X}_{\tau+1} \in B\right\},$$

as we wanted. □

5.5.3 Analysis on the Martingale Differences

In addition to the regenerative property of the inventory level, we also need to verify other conditions in Theorem 5.1 to prove Theorem 5.3. The first one is regarding the boundedness of the martingale differences.

Proposition 5.2 (Martingale difference upper bound). *There exists a constant B such that the martingale differences are uniformly upper-bounded*

$$\| d_{t,j} \|_{\infty} \leqslant B, \qquad \text{for all } j \in [p] \text{ and all } t \in [n].$$

The proof of this last boundedness property relies on the following intermediate step. This intermediate step essentially draws from Arlotto and Steele (2016, Lemma 14 and Lemma 15).

Proposition 5.3. *When the demand distribution satisfies the conditions in Definition 5.2, we have that for all $j \in [p]$ and all $t \in [n]$,*

$$\sup_{B \subset \mathbb{R}} \left| \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} - \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x'\right\} \right| \leqslant \bar{p} < 1,$$

with the constant $\bar{p}$ from Definition 5.2.

Proof. If we define $B_x \equiv \gamma_j(x) - B$, then from (5.12) and (5.15) we know that

$$\mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} = \mathbb{P}\left\{\xi_{1j} \in B_x\right\} = \int_{B_x} f_j(w)\, dw.$$

As a result, if we assume $x \leqslant x'$ and set $\epsilon = \gamma_j(x') - \gamma_j(x) \geqslant 0$, we obtain that

$$\mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} = \mathbb{P}\left\{\xi_{1j} - \epsilon \in B_x\right\} = \int_{B_x} f_j(w+\epsilon)\, dw.$$

Combining the last two identities, we further obtain that

$$\begin{aligned}
&\sup_{B \subset \mathbb{R}} \left| \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} - \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x'\right\} \right| \\
&= \sup_{B \subset \mathbb{R}} \left| \int_{B_x} f_j(w)\, dw - \int_{B_x} f_j(w+\epsilon)\, dw \right| = \left| \int_{B_x^*} f_j(w) - f_j(w+\epsilon)\, dw \right|,
\end{aligned}$$

where $B_x^* \equiv \{w : f_j(w) \geqslant f_j(w+\epsilon)\}$. Therefore, from Definition 5.2 we know that $B_x^* = [\widehat{w}, \bar{D}]$, so

$$\sup_{B\subset\mathbb{R}} \left| \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} - \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x'\right\} \right|$$

$$= \int_{\widehat{w}}^{\bar{D}} f_j(w) - f_j(w+\epsilon)\, dw = \mathbb{P}\left\{\widehat{w} \leqslant \xi_{1j} \leqslant \widehat{w} + \epsilon\right\}.$$

Since $\epsilon = \gamma_j(x') - \gamma_j(x) \leqslant S_j - s_j$, if $\widehat{w} + \epsilon \leqslant S_j$, then $\mathbb{P}\left\{\widehat{w} \leqslant \xi_{1j} \leqslant \widehat{w} + \epsilon\right\} \leqslant \mathbb{P}\left\{\xi_{1j} \leqslant S_j\right\}$, which is bounded by $\bar{p}$ by Definition 5.2. If $\widehat{w} + \epsilon > S_j$, then $\widehat{w} > s_j$, and so $\mathbb{P}\left\{\widehat{w} \leqslant \xi_{1j} \leqslant \widehat{w} + \epsilon\right\} \leqslant \mathbb{P}\left\{\xi_{1j} > s_j\right\}$, which is again bounded by $\bar{p}$ by Definition 5.2. Therefore, we obtain that

$$\sup_{B\subset\mathbb{R}} \left| \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x\right\} - \mathbb{P}\left\{X_{(t+1)j} \in B \;\middle|\; X_{tj} = x'\right\} \right| \leqslant \bar{p} < 1,$$

completing the proof. □

Now we are ready to prove Proposition 5.2.

Proof. (Proof of Proposition 5.2.) We first verify that the one-period cost of the dynamic inventory control problem is uniformly bounded. In fact, for the j-th decision maker, the cost she incurs during period t is given as

$$\mathcal{P}_{tj} = c_{o,j}(\gamma_j(X_{(t-1)j}) - X_{(t-1)j}) + L_j(\gamma_j(X_{(t-1)j}) - \xi_{tj}).$$

If we recall the ordering cost (5.13), the holding-and-shortage cost (5.14), and the order-up-to function (5.15), we obtain that

$$\mathcal{P}_{tj} = \begin{cases} c_{h,j}(X_{(t-1)j} - \xi_{tj}) & \text{if } s_j < X_{(t-1)j} \text{ and } \xi_{tj} \leqslant X_{(t-1)j} \\ c_{p,j}(\xi_{tj} - X_{(t-1)j}) & \text{if } s_j < X_{(t-1)j} < \xi_{tj} \\ K + c_{o,j}(S_j - X_{(t-1)j}) + c_{h,j}(S_j - \xi_{tj}) & \text{if } X_{(t-1)j} \leqslant s_j \text{ and } \xi_{tj} \leqslant S_j \\ K + c_{o,j}(S_j - X_{(t-1)j}) + c_{p,j}(\xi_{tj} - S_j) & \text{if } X_{(t-1)j} \leqslant s_j \text{ and } S_j < \xi_{tj}. \end{cases}$$

Since the demand ξ_{tj} is uniformly bounded within the interval $[0, \bar{D}]$, we can easily verify that the one-period cost satisfies that

$$0 \leqslant \mathcal{P}_{tj} \leqslant \max\left\{c_{h,j}S_j, c_{p,j}(\bar{D} - S_j), K + c_{o,j}(S_j - s_j) + c_{h,j}S_j, K + c_{o,j}(S_j - s_j) + c_{p,j}(\bar{D} - S_j)\right\} \equiv \bar{C}_j.$$

Then we set $\bar{C} = \max_{j\in[p]} \bar{C}_j$, and apply Arlotto and Steele (2016, Lemma 8) and Proposition 5.3 to obtain that the martingale differences satisfy that

$$\| d_{tj} \|_\infty \leqslant \frac{M\bar{C}}{\bar{p}} \equiv B \quad \text{for all } t \in [n] \text{ and all } j \in [p],$$

completing the proof. □

Next, we work towards an lower bound on the conditional second moment of martingale differences, $\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]$. This lower bound is summarized in the following lemma, whose proof relies on a coupling argument.

Lemma 5.4 (Martingale difference second moment lower bound). *There exists a constant $\ell_1 > 0$ such that the conditional second moment of the martingale differences are uniformly bounded from below, that is,*

$$\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] \geqslant \ell_1, \quad \text{for all } j \in [p] \text{ and all } t \in [n].$$

As a result, for any set $I = \{i+1, i+2, \ldots, i+q\}$, one has the following variance lower bound.

$$\operatorname{Var}\left\{\sum_{t\in I} d_{tj}\right\} \geqslant \ell_1 q.$$

Proof. Let $\{\widetilde{\xi}_{1j}, \ldots, \widetilde{\xi}_{nj}\}$ be an independent copy of the (actual) demands $\{\xi_{1j}, \ldots, \xi_{nj}\}$ and we use coupling to bound the conditional second moment $\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]$. Note that the inventory level $X_{(t-1)j}$ at the beginning of time period t is $\mathcal{F}_{t-1}$-measurable, i.e., it is totally determined by the demand realizations $\{\xi_{ij}\}_{i\in[t-1]}$. While the next period inventory level X_{tj} depends on $X_{(t-1)j}$ as well as the demand realization ξ_{tj} through the relation $X_{tj} = \gamma_j(X_{(t-1)j}) - \xi_{tj}$. Recall that for any two i.i.d. zero-mean random variables $Y \sim Z$, we have that

$$\operatorname{Var}\{Y\} = \mathbb{E}\left[Y^2\right] = \frac{1}{2}\mathbb{E}\left[(Y-Z)^2\right].$$

If we apply this to our martingale difference d_{tj} from (5.16), we would get

$$\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] = \frac{1}{2}\mathbb{E}\left[\left\{\widehat{v}_{(n-t)j}(\gamma_j(X_{(t-1)j}) - \xi_{tj}) - \widehat{v}_{(n-t)j}(\gamma_j(X_{(t-1)j}) - \widetilde{\xi}_{tj})\right\}^2 \,\middle|\, \mathcal{F}_{t-1}\right].$$

Next, we restrict ourselves on the event

$$G = G\left(X_{(t-1)j}\right) \equiv \left\{\omega :\ \xi_{tj}, \widetilde{\xi}_{tj} \in \left[\gamma_j(X_{(t-1)j}) - s_j, \gamma_j(X_{(t-1)j})\right]\right\}.$$

In fact, the choice of event G would guarantee that a new order will be placed for both of the two trajectories associated with $\{\xi_{tj}\}$ and $\{\widetilde{\xi}_{tj}\}$. To see this, one only needs to verify that the inventory level X_{tj} at the end of time period t is below the threshold s_j, that is,

$$\gamma_j(X_{(t-1)j}) - \xi_{tj} = X_{tj} \in [0, s_j].$$

The same holds for the other trajectory with $\{\widetilde{\xi}_{tj}\}$. As a result, conditioning on event G, the two trajectories coincide at time period $(t+1)$, and hence all the future randomnesses also become the same. This in turn gives us the lower bound

$$\begin{aligned}\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] \geqslant & \frac{1}{2}\mathbb{E}\left[(c_j + c_{h,j})^2 \left(\xi_{tj} - \widetilde{\xi}_{tj}\right)^2 \mathbb{1}\left\{G\right\} \,\middle|\, \mathcal{F}_{t-1}\right] \\ = & \frac{1}{2}(c_j + c_{h,j})^2 \iint\limits_{\Omega} (y-z)^2 f_j(y) f_j(z)\, dydz,\end{aligned}$$

where the integral region is $\Omega \equiv [\gamma_j(X_{(t-1)j}) - s_j, \gamma_j(X_{(t-1)j})]^2$, and f_j is the marginal density of the demand distribution faced by the j-th individual decision maker. The regularity condition (5.17) of demand distribution implies a strictly positive lower bound on the double integral

$$\iint\limits_{\Omega} (y-z)^2 f_j(y) f_j(z)\, dydz \geqslant \eta^2 \iint\limits_{\Omega} (y-z)^2\, dydz = \frac{1}{12}\eta^2 s_j^4,$$

hence

$$\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] \geqslant \frac{1}{24}(c_j + c_{h,j})^2 \eta_2^2 s_j^4 \geqslant \frac{\eta^2}{24} \min_{j\in[p]} \left\{(c_j + c_{h,j})^2 s_j^4\right\} \equiv \ell_1.$$

As a result, the variance is lower-bounded by

$$\mathrm{Var}\left\{\sum_{t\in I} d_{tj}\right\} = \sum_{t=1}^{q} \mathbb{E}\left[d_{(t+i)j}^2\right] = \sum_{t=1}^{q} \mathbb{E}\left[\mathbb{E}\left[d_{(t+i)j}^2 \,\middle|\, \mathcal{F}_{t+i-1}\right]\right] \geqslant \ell_1 q,$$

and the proof is complete. □

With all the above preparations, the proof of Theorem 5.3 becomes straightforward.

Proof. (Proof of Theorem 5.3.) We verify the conditions required in Theorem 5.1.

Condition (i) is verified by Proposition 5.2 and Lemma 5.4.

Condition (ii) is verified by Proposition 5.1 and Lemma 5.3. □

5.6 Application: Online Monotone Subsequence Selection from a Random Sample

In this section, we consider concrete application of the bootstrap procedure developed in Section 5.4. Specifically, we revisit the sequential monotone subsequence selection problem that was studied in Chapter 3.

5.6.1 Problem Description and Main Result

We remind the readers with the problem faced by a single decision maker described as follows. A sequence of i.i.d. random variables $W_1, \ldots, W_n$ from a continuous distribution are sequentially presented to the decision maker, who needs to select a (increasing) subsequence $W_{\tau_1}, \ldots, W_{\tau_\ell}$ that satisfies that

$$W_{\tau_1} \leqslant W_{\tau_2} \leqslant \cdots \leqslant W_{\tau_\ell},$$

where $\tau_1, \ldots, \tau_\ell$ are stopping times with respect to the filtration $\mathcal{F}_t = \sigma\{W_1, \ldots, W_t\}$, $t \in [n]$. Her objective is to maximize the expected length of the selected subsequence. This problem can be interpreted as a finite-horizon MDP, with the state X_t being the last selected value and the initial state being $X_0 = 0$. When the distribution F of the random variables is continuous, it is without loss of generality to assume that F is the uniform distribution on the unit interval. Moreover, the optimal policy is a time- and state-dependent threshold policy. Specifically, at any time t, let X_{t-1} be the last selected value and let W_t be the observed value. Then there is a threshold function $h_{n-t+1}(\cdot) : [0,1] \to [0,1]$ such that it is optimal to select W_t if and only if

$$X_{t-1} \leqslant W_t \leqslant h_{n-t+1}(X_{t-1}).$$

As a result, the state evolves accordingly and can be defined recursively as that for all $t \in [n]$,

$$X_t = \begin{cases} W_t & \text{if } W_t \in [X_{t-1}, h_{n-t+1}(X_{t-1})] \\ X_{t-1} & \text{otherwise.} \end{cases}$$

Then the total reward collected by this optimal policy is given as

$$R_n \equiv \sum_{t=1}^{n} \mathbb{1}\left\{W_t \in [X_{t-1}, h_{n-t+1}(X_{t-1})]\right\},$$

and one also has the optimal value function representation $v_n = \mathbb{E}[R_n]$.

The following upper bound on the length of the selection interval $[X_{t-1}, h_{n-t+1}(X_{t-1})]$ will be useful in our analysis.

Proposition 5.4 (Upper bound on selection interval length). *For any $t \in [n]$, we have that*

$$h_{n-t+1}(X_{t-1}) - X_{t-1} \leqslant \frac{4\log(n-t+1)+6}{\sqrt{2(n-t+1)}}. \tag{5.23}$$

Proof. We apply Theorem 3.1 to obtain a uniform upper bound on the difference $h_{n-t+1}(X_{(t-1)j}) - X_{(t-1)j}$ for all $j \in [p]$. In fact, we know from optimality that for all $j \in [p]$,

$$v_{n-t+1}(X_{(t-1)j}) \leqslant 1 + v_{n-t+1}(h_{n-t+1}(X_{(t-1)j})) \quad \text{for all } t \in [n].$$

By Theorem 3.1 we know that the left-hand side has lower bound

$$\sqrt{2(n-t+1)(1-X_{(t-1)j})} - 2(\log(n-t+1)+1) \leqslant v_{n-t+1}(X_{(t-1)j}),$$

and the right-hand side has upper bound

$$1 + v_{n-t+1}(h_{n-t+1}(X_{(t-1)j})) \leqslant 1 + \sqrt{2(n-t+1)(1-h_{n-t+1}(X_{(t-1)j}))}.$$

Putting these last three inequalities together, we further obtain that

$$\sqrt{2(n-t+1)(1-X_{(t-1)j})} - 2(\log(n-t+1)+1) \leqslant 1 + \sqrt{2(n-t+1)(1-h_{n-t+1}(X_{(t-1)j}))},$$

which in turn implies that

$$\begin{aligned} h_{n-t+1}(X_{(t-1)j}) - X_{(t-1)j} &\leqslant \frac{2\log(n-t+1)+3}{\sqrt{2(n-t+1)}} \left(\sqrt{1-X_{(t-1)j}} + \sqrt{1-h_{n-t+1}(X_{(t-1)j})}\right) \\ &\leqslant \frac{4\log(n-t+1)+6}{\sqrt{2(n-t+1)}}. \end{aligned}$$

□

Moreover, in this context, the Doob's martingale is simply $M_t = \mathbb{E}[R_n \mid \mathcal{F}_t]$ for $t = 0, 1, \ldots, n$, and hence the associated martingale differences are

$$d_t = M_t - M_{t-1} = \mathbb{E}[R_n \mid \mathcal{F}_t] - \mathbb{E}[R_n \mid \mathcal{F}_{t-1}], \quad \text{for all } t \in [n].$$

Assumption 5.1 (Bounded copulas). The joint distribution $F : [0,1]^p \to [0,1]$ satisfies that for any pair $i, j \in [p]$ and $i \neq j$, the bivariate distribution $F_{ij}(w_i, w_j)$ is smooth with density (copula) f_{ij}. Moreover, there exists constant $\bar{c}$ such that

$$f_{ij}(y,z) \leqslant \bar{c}, \quad \text{for all } i, j \in [p], i \neq j \text{ and all } (y,z) \in [0,1]^2.$$

In addition, for each $j \in [p]$, there are at most $\theta(n,p)$ coordinates that are correlated with j.

Remark 5.2. The sequential monotone subsequence selection problem is distribution invariant for any continuous distribution. As a result, whenever the marginal distributions are all continuous, it is without loss of generality to assume that all the marginal distributions are the uniform distribution on the unit interval.

When the joint distribution of the uncertainty satisfies Assumption 5.1, we are able to use bootstrap method to simultaneously construct confidence intervals for the total rewards collected by each decision maker, even when p is much larger than n. Such results are summarized in the following theorem.

Theorem 5.4. *Suppose that the distribution of uncertainty F satisfied Assumption 5.1. If*

$$p \leqslant \frac{\exp(n^{2a}\log^3 n)}{n^{3/4-3a/2}} \quad \textit{and} \quad \theta(n,p) \leqslant \frac{n^{1/8-3a/4}\log(np)}{\log^2 p \log^{3/2} n}, \tag{5.24}$$

then the simultaneous confidence intervals for the total rewards $R = (R_{n1}, \ldots, R_{np})$

$$\prod_{j=1}^{p}\left[\sqrt{2n} + \sum_{t=1}^{n}\widehat{d}_{tj} - \widehat{c}(\alpha)n^{1/4}, \sqrt{2n} + \sum_{t=1}^{n}\widehat{d}_{tj} + \widehat{c}(\alpha)n^{1/4}\right]$$

as constructed in Section 5.4 are asymptotic valid. That is,

$$\mathbb{P}\left\{\sqrt{2n} + \sum_{t=1}^{n}\widehat{d}_{tj} - \widehat{c}(\alpha)n^{1/4} \leqslant R_{nj} \leqslant \sqrt{2n} + \sum_{t=1}^{n}\widehat{d}_{tj} + \widehat{c}(\alpha)n^{1/4}\right\} \to (1-\alpha), \quad \textit{as } n \to \infty.$$

The proof of Theorem 5.4 relies on a detailed understanding of each individual MDP in order to verify the conditions required by Theorem 5.2. Such detailed properties of the individual MDP are stated in the following two subsections, after which we prove Theorem 5.4.

5.6.2 Control the Covariance Matrix

The variance of the total reward R_{nj} grows sublinearly, which requires a different normalization than the usual case. Specifically, it is well-known that $\mathrm{Var}\{R_{nj}\} \sim \sqrt{2n}/3$ as $n \to \infty$ (Arlotto et al., 2015, Theorem 1). Therefore, we scale up the martingale differences by $n^{1/4}$ to make the variance grow linearly in n. That is, we set

$$Z_{tj} \equiv n^{1/4} d_{tj} \quad \text{for all } t \in [n] \text{ and all } j \in [p]. \tag{5.25}$$

As a result, the diagonal entries of the covariance matrix

$$V \equiv \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[Z_t Z_t'\right] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\left[\boldsymbol{d}_t \boldsymbol{d}_t'\right] \tag{5.26}$$

would converge to constants as n goes to infinity. We also have the counterpart

$$V_n \equiv \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[Z_t Z_t' \mid \mathcal{F}_{t-1}\right] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\left[\boldsymbol{d}_t \boldsymbol{d}_t' \mid \mathcal{F}_{t-1}\right]. \tag{5.27}$$

Under suitable conditions of the uncertainty, we have the convergence in probability $V_n \xrightarrow{p} V$ as $n \to \infty$ with well controlled convergence rate. This result is summarized in the following proposition. We note that for the tail probability of $\left\| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right\|_\infty$, the Azuma-Hoeffding inequality would not be powerful enough due to the unusual normalization $n^{-1/2}$ on the right-hand side of (5.27). For instance, if we apply Azuma-Hoeffding inequality to the j-th diagonal entry of $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ by noticing $|\, d_{tj} \,| \leqslant 1$ for all $t \in [n]$, we would obtain that for any $\lambda > 0$,

$$\mathbb{P}\left\{ \left| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right|_{jj} > \lambda \right\} = \mathbb{P}\left\{ \left| \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]\right) \right| > \lambda \right\} \leqslant 2\exp(-\lambda^2/2),$$

which is too loose for our purpose.

Proposition 5.5 (Convergence in probability of the covariance matrix). *If the uncertainty distribution F satisfies Assumption 5.1, then there exists a constant $c_1 > 0$ depends only on $\bar{c}$, such that for any fixed $a \in (0, 1/6)$, with probability at least $1 - 3p^2\exp(-n^{2a})$, we have that*

$$\max_{i,j\in[p]} |\, V_{n,ij} - V_{ij} \,| \leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}},$$

as well as that

$$\max_{i,j\in[p]} \left| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right| \leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}.$$

As a result, we also have that

$$\mathbb{E}\left[\max_{i,j\in[p]} |\, V_n - V \,|_{ij} \right] \leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}} + 3p^2 n^{1/2}\exp(-n^{2a}).$$

In order to prove Proposition 5.5, we need some intermediate results. Essentially, we are interested in the convergences in probability

$$\| V_n - V \|_\infty \xrightarrow{p} 0 \quad \text{and} \quad \left\| \frac{1}{n} \sum_{t=1}^{n} Z_t Z_t' - V_n \right\|_\infty \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

To facilitate our analysis, we define another sequence of martingale differences $\Delta_t \in \mathbb{R}^{p \times p}$ for all $t \in [n]$ as

$$\Delta_t \equiv \mathbb{E}\left[V_n \mid \mathcal{F}_t \right] - \mathbb{E}\left[V_n \mid \mathcal{F}_{t-1} \right], \tag{5.28}$$

so that we have the representation

$$V_n - V = \sum_{t=1}^{n} \Delta_t. \tag{5.29}$$

The proof of Proposition 5.5 requires two intermediate technical propositions. The first technical proposition is the following.

Proposition 5.6. *Suppose the uncertainty distribution F satisfies Assumption 5.1, then for all $j \in [p]$, the j-th diagonal entry of $| V_n - V |$ has the tail probability bound that for any $\lambda > 0$ and $\nu \geqslant \sqrt{2}(6 + 4 \log n/3)^2/\sqrt{n}$,*

$$\mathbb{P}\left\{ | V_n - V |_{jj} > \lambda \right\} \leqslant 2 \exp\left(-\frac{\lambda^2}{2\nu} \right) + \exp\left(- \frac{\left(\frac{n\nu}{(6+4\log n/3)^2} - \sqrt{2n} \right)^2}{2n} \right).$$

For all $i, j \in [p]$ with $i \neq j$, the (i,j)-entry of $| V_n - V |$ has the tail probability bound that for any $\lambda > 0$ and $\nu \geqslant 8\sqrt{2}(\bar{c}+1)^2(2\log n + 3)^6/(9\sqrt{n})$,

$$\mathbb{P}\left\{ | V_n - V |_{ij} > \lambda \right\} \leqslant 2 \exp\left(-\frac{\lambda^2}{2\nu} \right) + \exp\left(- \frac{\left(\frac{9n\nu}{4(\bar{c}+1)^2(2\log n+3)^6} - 2\sqrt{2n} \right)^2}{2n} \right).$$

Proof. Step 1. Diagonal entries of $V_n - V$.

From Arlotto et al. (2015, Lemma 24) we know that

$$| (\Delta_t)_{jj} | \leqslant \frac{6 + \frac{4}{3}\log n}{\sqrt{n}} \mathbb{1}\left\{ G_t^j \right\}, \tag{5.30}$$

where the event G_t^j is when a selection happens, that is,

$$G_t^j \equiv \left\{\omega : \ W_{tj} \in \left[X_{(t-1)j}, h_{n-t+1}(X_{(t-1)j})\right]\right\}.$$

We know from (5.29) that the tail probability of the j-th diagonal entry of $V_n - V$ can be represented as

$$\begin{aligned}\mathbb{P}\left\{|\, V_n - V\,|_{jj} > \lambda\right\} =& \mathbb{P}\left\{\left|\sum_{t=1}^n (\Delta_t)_{jj}\right| > \lambda, \sum_{t=1}^n (\Delta_t)_{jj}^2 \leqslant \nu\right\} + \mathbb{P}\left\{\left|\sum_{t=1}^n (\Delta_t)_{jj}\right| > \lambda, \sum_{t=1}^n (\Delta_t)_{jj}^2 > \nu\right\} \\ \leqslant& \mathbb{P}\left\{\left|\sum_{t=1}^n (\Delta_t)_{jj}\right| > \lambda, \sum_{t=1}^n (\Delta_t)_{jj}^2 \leqslant \nu\right\} + \mathbb{P}\left\{\sum_{t=1}^n (\Delta_t)_{jj}^2 > \nu\right\}. \end{aligned} \tag{5.31}$$

To bound the first probability, we apply Belloni and Oliveira (2017, Lemma 9) to obtain that

$$\mathbb{P}\left\{\left|\sum_{t=1}^n (\Delta_t)_{jj}\right| > \lambda, \sum_{t=1}^n (\Delta_t)_{jj}^2 \leqslant \nu\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2\nu}\right). \tag{5.32}$$

On the other hand, we know from (5.30) that the second probability on the right-hand side of (5.31) is bounded by

$$\mathbb{P}\left\{\sum_{t=1}^n (\Delta_t)_{jj}^2 > \nu\right\} \leqslant \mathbb{P}\left\{\sum_{t=1}^n \mathbb{1}\left\{G_t^j\right\} > \frac{n\nu}{(6+4\log n/3)^2}\right\}.$$

Then we apply Azuma-Hoeffding inequality to obtain that

$$\begin{aligned}&\mathbb{P}\left\{\sum_{t=1}^n \mathbb{1}\left\{G_t^j\right\} > \frac{n\nu}{(6+4\log n/3)^2}\right\} \\ =&\mathbb{P}\left\{\sum_{t=1}^n \left\{\mathbb{1}\left\{G_t^j\right\} - \mathbb{E}\left[\mathbb{1}\left\{G_t^j\right\}\right]\right\} > \frac{n\nu}{(6+4\log n/3)^2} - \sum_{t=1}^n \mathbb{E}\left[\mathbb{1}\left\{G_t^j\right\}\right]\right\} \\ \leqslant& \exp\left(-\frac{\left(\frac{n\nu}{(6+4\log n/3)^2} - \sum_{t=1}^n \mathbb{E}\left[\mathbb{1}\left\{G_t^j\right\}\right]\right)^2}{2n}\right).\end{aligned}$$

Another application of Arlotto et al. (2015, Theorem 1, Equation (5)) yields that $\sum_{t=1}^n \mathbb{E}\left[\mathbb{1}\left\{G_t^j\right\}\right] \leqslant \sqrt{2n}$, so if we choose ν such that

$$\frac{n\nu}{(6+4\log n/3)^2} \geqslant \sqrt{2n},$$

then we obtain an upper bound on the second piece of the right-hand side of (5.31), that is,

$$\mathbb{P}\left\{\sum_{t=1}^{n}(\Delta_t)_{jj}^2 > \nu\right\} \leqslant \exp\left(-\frac{\left(\frac{n\nu}{(6+4\log n/3)^2} - \sqrt{2n}\right)^2}{2n}\right). \tag{5.33}$$

Combining (5.32) and (5.33) together, we finally obtain the tail probability bound on the j-th diagonal entry of $|\, V_n - V \,|$ given as

$$\mathbb{P}\left\{|\, V_n - V \,|_{jj} > \lambda\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2\nu}\right) + \exp\left(-\frac{\left(\frac{n\nu}{(6+4\log n/3)^2} - \sqrt{2n}\right)^2}{2n}\right),$$

for all $\lambda > 0$ and all $\nu \geqslant \sqrt{2/n}(6 + 4\log n/3)^2$.

Step 2. Off-diagonal entries of $V_n - V$.

For $i, j \in [p]$ and $i \neq j$, the (i,j)-element of the martingale difference Δ_t from (5.28) is simply

$$\begin{aligned}(\Delta_t)_{ij} =& \mathbb{E}\left[(V_n)_{ij} \mid \mathcal{F}_t\right] - \mathbb{E}\left[(V_n)_{ij} \mid \mathcal{F}_{t-1}\right] \\ =& \frac{1}{\sqrt{n}}\sum_{k=t+1}^{n}\mathbb{E}\left[d_{ki}d_{kj} \mid \mathcal{F}_t\right] - \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=t+1}^{n}\mathbb{E}\left[d_{ki}d_{kj} \mid \mathcal{F}_t\right] \,\middle|\, \mathcal{F}_{t-1}\right].\end{aligned} \tag{5.34}$$

If we introduce the shorthand $\omega_{n-t}(X_{ti}, X_{tj}) \equiv \sum_{k=t+1}^{n}\mathbb{E}\left[d_{ki}d_{kj} \mid \mathcal{F}_t\right]$, then this last equation can be re-written as

$$\sqrt{n}(\Delta_t)_{ij} = \omega_{n-t}(X_{ti}, X_{tj}) - \mathbb{E}\left[\omega_{n-t}(X_{ti}, X_{tj}) \mid \mathcal{F}_{t-1}\right]. \tag{5.35}$$

Each one of the original martingale differences, d_{ti}, can be represented as the sum of two quantities depending on what happens at period t. Specifically, one has that $d_{ti} = A_{ti} + B_{ti}$ (c.f. Arlotto et al., 2015, Equation (45)), where

$$\begin{aligned}A_{ti} \equiv& (1 + v_{n-t}(W_{ti}) - v_{n-t}(X_{(t-1)i}))\mathbb{1}\left\{W_{ti} \in [X_{(t-1)i}, h_{n-t+1}(X_{(t-1)i})]\right\}, \\ B_{ti} \equiv& v_{n-t}(X_{(t-1)i}) - v_{n-t+1}(X_{(t-1)i}).\end{aligned}$$

Since d_{ti} is a martingale difference, we have the identity $\mathbb{E}\left[A_{ti} \mid \mathcal{F}_{t-1}\right] = B_{ti}$. As a result,

$$\mathbb{E}\left[d_{ti}d_{tj} \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[A_{ti}A_{tj} \mid \mathcal{F}_{t-1}\right] - B_{ti}B_{tj}.$$

The $\mathcal{F}_{t-1}$-measurable quantity $B_{ti}B_{tj}$ is easy to bound since for each one of the two pieces, we

have that

$$\begin{aligned}
|\, B_{ti} \,| &= \left|\, v_{n-t}(X_{(t-1)i}) - v_{n-t+1}(X_{(t-1)i}) \,\right| \\
&= \left| \int_{X_{(t-1)i}}^{h_{n-t+1}(X_{(t-1)i})} \left\{1 + v_{n-t}(w) - v_{n-t}(X_{(t-1)i})\right\} dw \right| \\
&\leqslant h_{n-t+1}(X_{(t-1)i}) - X_{(t-1)i},
\end{aligned}$$

and from (5.23) we know that this last right-hand side is at most $\frac{4\log(n-t+1)+6}{\sqrt{2(n-t+1)}}$. As a result,

$$|\, B_{ti}B_{tj} \,| \leqslant \frac{2\left(2\log(n-t+1)+3\right)^2}{(n-t+1)}. \tag{5.36}$$

On the other hand, we have the integral representation

$$\begin{aligned}
&\mathbb{E}\left[A_{ti}A_{tj} \mid \mathcal{F}_{t-1}\right] \\
&= \iint\limits_{\Lambda} \left\{1 + v_{n-t}(w_1) - v_{n-t}(X_{(t-1)i})\right\} \left\{1 + v_{n-t}(w_2) - v_{n-t}(X_{(t-1)j})\right\} f_{ij}(w_1, w_2)\, dw_1 dw_2,
\end{aligned}$$

where the double integral is over the region $\Lambda \equiv \left[X_{(t-1)i}, h_{n-t+1}(X_{(t-1)i})\right] \times \left[X_{(t-1)j}, h_{n-t+1}(X_{(t-1)j})\right]$, and $f_{ij}(\cdot,\cdot)$ is the density of the joint uncertainty faced by the i-th and the j-th decision makers.

Then by Assumption 5.1 we know that

$$\mathbb{E}\left[A_{ti}A_{tj} \mid \mathcal{F}_{t-1}\right] \leqslant \overline{c}\left(h_{n-t+1}(X_{(t-1)i}) - X_{(t-1)i}\right)\left(h_{n-t+1}(X_{(t-1)j}) - X_{(t-1)j}\right),$$

and if we apply (5.23) once again, we would obtain that

$$\mathbb{E}\left[A_{ti}A_{tj} \mid \mathcal{F}_{t-1}\right] \leqslant \frac{2\overline{c}\left(2\log(n-t+1)+3\right)^2}{(n-t+1)}. \tag{5.37}$$

Finally, we assemble the estimations (5.37) and (5.36) together to obtain that

$$\begin{aligned}
|\, \omega_{n-t}(X_{ti}, X_{tj}) \,| = \left| \sum_{k=t+1}^{n} \mathbb{E}\left[d_{ki}d_{kj} \mid \mathcal{F}_t\right] \right| &= \left| \sum_{k=t+1}^{n} \mathbb{E}\left[A_{ki}A_{kj} \mid \mathcal{F}_t\right] - \sum_{k=t+1}^{n} \mathbb{E}\left[B_{ki}B_{kj} \mid \mathcal{F}_t\right] \right| \\
&= \sum_{k=t+1}^{n} \frac{2(\overline{c}+1)\left(2\log(n-k+1)+3\right)^2}{(n-k+1)} \\
&\leqslant \frac{2(\overline{c}+1)\left(2\log(n-t)+3\right)^3}{3}.
\end{aligned} \tag{5.38}$$

Consider the following three disjoint events that describe what could happen at time period t,

$$G_t^{ij} \equiv \left\{\omega : \ W_{ti} \in [X_{(t-1)i}, h_{n-t+1}(X_{(t-1)i})], W_{tj} \in [X_{(t-1)j}, h_{n-t+1}(X_{(t-1)j})]\right\}$$

$$G_t^{i} \equiv \left\{\omega : \ W_{ti} \in [X_{(t-1)i}, h_{n-t+1}(X_{(t-1)i})], W_{tj} \notin [X_{(t-1)j}, h_{n-t+1}(X_{(t-1)j})]\right\}$$

$$G_t^{j} \equiv \left\{\omega : \ W_{ti} \notin [X_{(t-1)i}, h_{n-t+1}(X_{(t-1)i})], W_{tj} \in [X_{(t-1)j}, h_{n-t+1}(X_{(t-1)j})]\right\}.$$

In words, the event G_t^{ij} is that both the i-th decision maker and the j-th decision maker make a selection in time period t; the event G_t^i is that the i-th decision maker makes a selection in time period t while the j-th decision maker does not; the event G_t^i is that the j-th decision maker makes a selection in time period t while the i-th decision maker does not. Then we can represent $\omega_{n-t}(X_{ti}, X_{tj})$ in terms of the three events as

$$\begin{aligned}\omega_{n-t}(X_{ti}, X_{tj}) = \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j}) &+ \{\omega_{n-t}(W_{ti}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{ij}\right\} \\ &+ \{\omega_{n-t}(W_{ti}, X_{(t-1)j}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{i}\right\} \\ &+ \{\omega_{n-t}(X_{(t-1)i}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{j}\right\}.\end{aligned}$$

Since the first term on this last right-hand side is $\mathcal{F}_{t-1}$-measurable, applying this last equation to (5.35) implies that

$$\begin{aligned}\sqrt{n}(\Delta_t)_{ij} =& \{\omega_{n-t}(W_{ti}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{ij}\right\} \\ &- \mathbb{E}\left[\{\omega_{n-t}(W_{ti}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{ij}\right\} \mid \mathcal{F}_{t-1}\right] \\ &+ \{\omega_{n-t}(W_{ti}, X_{(t-1)j}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{i}\right\} \\ &- \mathbb{E}\left[\{\omega_{n-t}(W_{ti}, X_{(t-1)j}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{i}\right\} \mid \mathcal{F}_{t-1}\right] \\ &+ \{\omega_{n-t}(X_{(t-1)i}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{j}\right\} \\ &- \mathbb{E}\left[\{\omega_{n-t}(X_{(t-1)i}, W_{tj}) - \omega_{n-t}(X_{(t-t)i}, X_{(t-1)j})\}\mathbb{1}\left\{G_t^{j}\right\} \mid \mathcal{F}_{t-1}\right].\end{aligned}$$

Next, if we apply the uniform bound (5.38), we would obtain that

$$|\,(\Delta_t)_{ij}\,| \leqslant \frac{2(\overline{c}+1)\left(2\log(n-t)+3\right)^3}{3\sqrt{n}}\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\}, \tag{5.39}$$

and so the same argument from *Step 1* applies here. Basically, from (5.29) we know that the tail

probability of $|V_n - V|_{ij}$ satisfies that

$$\mathbb{P}\left\{|V_n - V|_{ij} > \lambda\right\} = \mathbb{P}\left\{\left|\sum_{t=1}^{n}(\Delta_t)_{ij}\right| > \lambda\right\}$$

$$=\mathbb{P}\left\{\left|\sum_{t=1}^{n}(\Delta_t)_{ij}\right| > \lambda, \sum_{t=1}^{n}(\Delta_t)_{ij}^2 \leqslant \nu\right\} + \mathbb{P}\left\{\left|\sum_{t=1}^{n}(\Delta_t)_{ij}\right| > \lambda, \sum_{t=1}^{n}(\Delta_t)_{ij}^2 > \nu\right\}$$

$$\leqslant\mathbb{P}\left\{\left|\sum_{t=1}^{n}(\Delta_t)_{ij}\right| > \lambda, \sum_{t=1}^{n}(\Delta_t)_{ij}^2 \leqslant \nu\right\} + \mathbb{P}\left\{\sum_{t=1}^{n}(\Delta_t)_{ij}^2 > \nu\right\}.$$

We apply Belloni and Oliveira (2017, Lemma 9) to the first term of this last right-hand side to obtain that

$$\mathbb{P}\left\{\left|\sum_{t=1}^{n}(\Delta_t)_{ij}\right| > \lambda, \sum_{t=1}^{n}(\Delta_t)_{ij}^2 \leqslant \nu\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2\nu}\right). \tag{5.40}$$

On the other hand, we apply (5.39) to obtain that

$$\mathbb{P}\left\{\sum_{t=1}^{n}(\Delta_t)_{ij}^2 > \nu\right\} \leqslant \mathbb{P}\left\{\sum_{t=1}^{n}\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\} > \frac{9n\nu}{4(\bar{c}+1)^2\left(2\log n+3\right)^6}\right\}.$$

Then we apply Azuma-Hoeffding inequality to obtain that

$$\mathbb{P}\left\{\sum_{t=1}^{n}\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\} > \frac{9n\nu}{4(\bar{c}+1)^2\left(2\log n+3\right)^6}\right\}$$

$$=\mathbb{P}\left\{\sum_{t=1}^{n}\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\} - \mathbb{E}\left[\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\}\right] > \frac{9n\nu}{4(\bar{c}+1)^2\left(2\log n+3\right)^6} - \sum_{t=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\}\right]\right\}$$

$$\leqslant \exp\left(-\frac{\left(\frac{9n\nu}{4(\bar{c}+1)^2(2\log n+3)^6} - \sum_{t=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\}\right]\right)^2}{2n}\right).$$

Since one always has

$$\mathbb{1}\left\{G_t^{ij} \cup G_t^i \cup G_t^j\right\} \leqslant \mathbb{1}\left\{G_t^i\right\} + \mathbb{1}\left\{G_t^j\right\},$$

and

$$\sum_{t=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{G_t^i\right\}\right] \leqslant \sqrt{2n},$$

so if we choose ν such that

$$\frac{9n\nu}{4(\bar{c}+1)^2\left(2\log n+3\right)^6} \geqslant 2\sqrt{2n},$$

then we obtain an upper bound

$$\mathbb{P}\left\{\sum_{t=1}^{n}(\Delta_t)_{ij}^2 > \nu\right\} \leqslant \mathbb{P}\left\{\sum_{t=1}^{n}\mathbb{1}\left\{G_t^{ij}\cup G_t^i\cup G_t^j\right\} > \frac{9n\nu}{4(\bar{c}+1)^2\left(2\log n+3\right)^6}\right\}$$

$$\leqslant \exp\left(-\frac{\left(\frac{9n\nu}{4(\bar{c}+1)^2(2\log n+3)^6}-2\sqrt{2n}\right)^2}{2n}\right). \tag{5.41}$$

Putting (5.40) and (5.41) together, we obtain that

$$\mathbb{P}\left\{|\,V_n - V\,|_{ij} > \lambda\right\} \leqslant 2\exp(-\frac{\lambda^2}{2\nu}) + \exp\left(-\frac{\left(\frac{9n\nu}{4(\bar{c}+1)^2(2\log n+3)^6}-2\sqrt{2n}\right)^2}{2n}\right),$$

for all $\lambda > 0$ and all $\nu \geqslant 8\sqrt{2}(\bar{c}+1)^2(2\log n+3)^6/(9\sqrt{n})$. □

The second intermediate technical proposition is the following.

Proposition 5.7. *For all $\lambda > 0$ and all $\nu \geqslant \frac{\sqrt{2n}+2(\log n+1)}{3n}$, the j-th diagonal entry of the matrix $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ has the tail probability bound*

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n\right|_{jj} > \lambda\right\} \leqslant 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2-\exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right] + \exp\left(-\frac{\left(n\nu-\frac{\sqrt{2n}}{3}-\frac{2}{3}(1+\log n)\right)^2}{2n\left(6+\frac{4}{3}\log n\right)}\right).$$

For any pair (i,j) with $i \neq j$, the (i,j)-entry of the matrix $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ has the tail probability bound

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n\right|_{ij} > \lambda\right\} \leqslant 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2-\exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right] + \exp\left(-\frac{\left(n\nu-\frac{\sqrt{2n}}{3}-\frac{2}{3}(1+\log n)\right)^2}{2n\left(6+\frac{4}{3}\log n\right)}\right).$$

Proof. Step 1. Diagonal entries of $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$.

From (5.25) and (5.27), we know that the j-th diagonal entry of the matrix $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ is given by

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right],$$

so with the shorthand $\eta_{tj} \equiv d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]$ for all $t \in [n]$ and all $j \in [p]$, one can easily verify that $\{\eta_{tj}\}_{t=1}^n$ is a sequence of martingale differences as $\mathbb{E}\left[\eta_{tj} \mid \mathcal{F}_{t-1}\right] = 0$. Moreover, for all $j \in [p]$, we also have the tail probability bound

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]\right| > \lambda\right\} = \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \eta_{tj}\right| > \lambda\right\} \\
=&\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \eta_{tj}\right| > \lambda,\ \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} + \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \eta_{tj}\right| > \lambda,\ \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \\
\leqslant&\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \eta_{tj}\right| > \lambda,\ \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} + \mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\}.
\end{aligned}$$

As for the first piece of this last right-hand side, note that $|\,\eta_{tj}\,| \leqslant 1$, so we apply Belloni and Oliveira (2017, Lemma 10) to obtain that

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \eta_{tj}\right| > \lambda,\ \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} \\
=&\mathbb{P}\left\{\left|\sum_{t=1}^n \eta_{tj}\right| > \sqrt{n}\lambda,\ \sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant n\nu\right\} \leqslant 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2 - \exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right]. \qquad (5.42)
\end{aligned}$$

While for the second piece, from the definition $\eta_{tj} = d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]$ we know that

$$\mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[d_{tj}^4 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]^2 \leqslant \mathbb{E}\left[d_{tj}^4 \mid \mathcal{F}_{t-1}\right],$$

and since $|\,d_{tj}\,| \leqslant 1$ (see, e.g., Arlotto et al., 2015, Equation (43)), we further have that $\mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]$. As a result,

$$\begin{aligned}
\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \leqslant&\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \\
=&\mathbb{P}\left\{\sum_{t=1}^n \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2\right] > n\nu - \sum_{t=1}^n \mathbb{E}\left[d_{tj}^2\right]\right\}.
\end{aligned}$$

In the rightmost probability of this last equation, we can once again represent the zero-mean quantity as a sum of martingale differences. Recall the two matrices (5.26) and (5.27), we know that

$$\sum_{t=1}^n \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2\right] = \sqrt{n}(V_n - V)_{jj} = \sum_{t=1}^n \sqrt{n}(\Delta_t)_{jj}, \qquad (5.43)$$

and from (5.30) we further obtain that each martingale difference is bounded by

$$\sqrt{n}\,|\,(\Delta_t)_{jj}\,| \leqslant \left(6 + \frac{4}{3}\log n\right) \mathbb{1}\left\{G_t^j\right\}.$$

Then we apply Azuma-Hoeffding inequality to the right tail of the sum of the martingale differences in (5.43) to obtain that

$$\mathbb{P}\left\{\sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2\right] > n\nu - \sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right]\right\} \leqslant \exp\left(-\frac{\left(n\nu - \sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right]\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right).$$

By Arlotto et al. (2015, Theorem 1 and Proposition 20) we know that

$$\sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right] = \mathrm{Var}\left\{\sum_{t=1}^{n} d_{tj}\right\} \leqslant \frac{\sqrt{2n}}{3} + \frac{2}{3}(1 + \log n), \tag{5.44}$$

which in turn implies that if $n\nu \geqslant \frac{\sqrt{2n}}{3} + \frac{2}{3}(1 + \log n)$, then

$$\mathbb{P}\left\{\sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2\right] > n\nu - \sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right]\right\} \leqslant \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1 + \log n)\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right), \tag{5.45}$$

and so we finally obtain that

$$\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \leqslant \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1 + \log n)\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right). \tag{5.46}$$

Putting (5.42) and (5.46) together, we obtain the final bound that for all $t > 0$ and all $\nu \geqslant \frac{\sqrt{2n}+2(1+\log n)}{3n}$,

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{tj}^2 - \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right]\right| > \lambda\right\} \\
\leqslant &2\exp\left[-\frac{\lambda^2}{2\nu}\left(2 - \exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right] + \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1 + \log n)\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right).
\end{aligned}$$

Step 2. Off diagonal entries of $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$.

The (i,j)-element of the matrix $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ is given by

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} d_{ti} d_{tj} - \mathbb{E}\left[d_{ti} d_{tj} \mid \mathcal{F}_{t-1}\right],$$

so with the shorthand $\eta_{tij} \equiv d_{ti} d_{tj} - \mathbb{E}\left[d_{ti} d_{tj} \mid \mathcal{F}_{t-1}\right]$ for all $t \in [n]$ and all $i, j \in [p]$ with $i \neq j$, one can easily verify that $\{\eta_{tij}\}_{t\in[n]}$ is a sequence of martingale differences as $\mathbb{E}\left[\eta_{tij} \mid \mathcal{F}_{t-1}\right] = 0$. Moreover, for all $i, j \in [p]$ with $i \neq j$, we also have the tail probability bound

$$\begin{aligned}
&\mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} d_{ti} d_{tj} - \mathbb{E}\left[d_{ti} d_{tj} \mid \mathcal{F}_{t-1}\right] \right| > \lambda\right\} = \mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{tij} \right| > \lambda\right\} \\
=&\mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{tij} \right| > \lambda,\ \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} + \mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{tij} \right| > \lambda,\ \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \\
\leqslant&\mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{tij} \right| > \lambda,\ \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} + \mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\}.
\end{aligned}$$

As for the first piece of this last right-hand side, note that $|\,\eta_{tij}\,| \leqslant 1$, so we apply Belloni and Oliveira (2017, Lemma 10) to obtain that

$$\mathbb{P}\left\{\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_{tij} \right| > \lambda,\ \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \nu\right\} \leqslant 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2 - \exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right]. \tag{5.47}$$

While for the second piece, from the definition $\eta_{tij} = d_{ti} d_{tj} - \mathbb{E}\left[d_{ti} d_{tj} \mid \mathcal{F}_{t-1}\right]$ and the fact that $|\,d_{ti}\,| \leqslant 1$ (see, e.g., Arlotto et al., 2015, Equation (43)), we have the crude bound

$$\mathbb{E}\left[\eta_{tj}^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[d_{ti}^2 d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{ti} d_{tj} \mid \mathcal{F}_{t-1}\right]^2 \leqslant \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right].$$

As a result,

$$\begin{aligned}
\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \leqslant&\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \\
=&\mathbb{P}\left\{\sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[d_{tj}^2\right] > n\nu - \sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right]\right\}.
\end{aligned}$$

From (5.45), we know that

$$\mathbb{P}\left\{\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\eta_{tij}^2 \mid \mathcal{F}_{t-1}\right] > \nu\right\} \leqslant \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1+\log n)\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right). \tag{5.48}$$

Lastly, if we put (5.47) and (5.48) together, we obtain that the (i,j)-element of $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$ has the tail bound

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n\right|_{ij} > \lambda\right\} =& \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{ti}d_{tj} - \mathbb{E}\left[d_{ti}d_{tj} \mid \mathcal{F}_{t-1}\right]\right| > \lambda\right\} \\ \leqslant& 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2-\exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right] + \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1+\log n)\right)^2}{2n\left(6+\frac{4}{3}\log n\right)}\right),\end{aligned}$$

so the proof is complete. □

Now we are ready to prove Proposition 5.5.

Proof. (Proof of Proposition 5.5.) We start with the convergence $\| V_n - V \|_\infty \xrightarrow{p} 0$. From Proposition 5.6 we know that for the j-th diagonal entry of $V_n - V$, its tail probability satisfies that

$$\mathbb{P}\left\{|\, V_n - V\,|_{jj} > \lambda\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2\nu}\right) + \exp\left(-\frac{\left(\frac{n\nu}{(6+4\log n/3)^2} - \sqrt{2n}\right)^2}{2n}\right),$$

so if we fix some $a \in (0, 1/6)$ and choose

$$\lambda = c_1 n^{3a/2-1/4}\log^3 n \quad \text{and} \quad \nu = \frac{c_1 n^{a-1/2}\log^6 n}{2}, \tag{5.49}$$

then we can choose proper constant c_1 to obtain that

$$\mathbb{P}\left\{|\, V_n - V\,|_{jj} > \frac{c_1\log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant 3\exp(-n^{2a}).$$

Similarly, for the (i,j)-entry of $V_n - V$ with $i \neq j$, we know from Proposition 5.6 that its tail probability satisfies that

$$\mathbb{P}\left\{|\, V_n - V\,|_{ij} > \lambda\right\} \leqslant 2\exp\left(-\frac{\lambda^2}{2\nu}\right) + \exp\left(-\frac{\left(\frac{9n\nu}{4(\bar{c}+1)^2(2\log n+3)^6} - 2\sqrt{2n}\right)^2}{2n}\right).$$

With the same choice of λ and ν, we can choose proper constant c_1 that depends only on $\bar{c}$ to obtain that

$$\mathbb{P}\left\{|\, V_n - V\,|_{ij} > \frac{c_1\log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant 3\exp(-n^{2a}).$$

Therefore, we have the union bound that

$$\mathbb{P}\left\{\| V_n - V \|_\infty > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant \sum_{j=1}^{n} \mathbb{P}\left\{| V_n - V |_{jj} > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} + \sum_{i\neq j} \mathbb{P}\left\{| V_n - V |_{ij} > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\}$$

$$\leqslant 3p^2 \exp(-n^{2a}).$$

Next, we prove that $\left\| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right\|_\infty \xrightarrow{p} 0$. From Proposition 5.7 we know that for the (i,j)-entry of $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n$, its tail probability satisfies that

$$\mathbb{P}\left\{\left| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right|_{ij} > \lambda\right\} \leqslant 2\exp\left[-\frac{\lambda^2}{2\nu}\left(2 - \exp(\frac{\lambda}{\sqrt{n}\nu})\right)\right] + \exp\left(-\frac{\left(n\nu - \frac{\sqrt{2n}}{3} - \frac{2}{3}(1+\log n)\right)^2}{2n\left(6 + \frac{4}{3}\log n\right)}\right).$$

If we choose the same λ and ν as in (5.49), then with proper choice of c_1 we obtain that

$$\mathbb{P}\left\{\left| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right|_{ij} > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant 3\exp(-n^{2a}).$$

As a result, we have the union bound

$$\mathbb{P}\left\{\left\| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right\|_\infty > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant \sum_{i,j\in[p]} \mathbb{P}\left\{\left| \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' - V_n \right|_{ij} > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant 3p^2\exp(-n^{2a}).$$

Lastly, we prove the upper bound on $\mathbb{E}[\| V_n - V \|_\infty]$. Since we have already proved that

$$\mathbb{P}\left\{\| V_n - V \|_\infty > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\} \leqslant 3p^2\exp(-n^{2a}),$$

we can bound $\mathbb{E}[\| V_n - V \|_\infty]$ from above by

$$\mathbb{E}[\| V_n - V \|_\infty] \leqslant \mathbb{P}\left\{\| V_n - V \|_\infty \leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\}\frac{\log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}} + \mathbb{P}\left\{\| V_n - V \|_\infty > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\}(\| V_n \|_\infty + \| V \|_\infty)$$

$$\leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}} + \mathbb{P}\left\{\| V_n - V \|_\infty > \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}\right\}(\| V_n \|_\infty + \| V \|_\infty).$$

Recall the matrices V and V_n from (5.26) and (5.27) receptively, and recall that $| d_{tj} | \leqslant 1$, we have the crude upper bound $\| V \|_\infty \leqslant \sqrt{n}$ and $\| V_n \|_\infty \leqslant \sqrt{n}$. Therefore, with some algebra, we finally obtain that

$$\mathbb{E}[\| V_n - V \|_\infty] \leqslant \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}} + 3p^2 n^{1/2}\exp(-n^{2a}),$$

and the proof is complete. □

5.6.3 Verification of Other Conditions

We start with the boundedness condition on $\| V \|_\infty$ in Condition (ii) of Theorem 5.2.

Proposition 5.8 (Boundedness of covariance matrix entries). *The covariance matrix V defined in* (5.26) *has the uniform bound on all of its entries given as*

$$\max_{i,j\in[p]} | V_{ij} | \leqslant \frac{\sqrt{2}}{3} + \frac{2(1+\log n)}{3\sqrt{n}}.$$

Proof. We start with the diagonal entries of V. From the definition of V in (5.26) and the upper bound (5.44), we know that for all $j \in [p]$,

$$V_{jj} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{E}\left[d_{tj}^2\right] = \frac{1}{\sqrt{n}} \mathbb{E}\left[\left(\sum_{t=1}^{n} d_{tj}\right)^2\right] = \frac{1}{\sqrt{n}} \mathrm{Var}\left\{R_{nj}\right\} \leqslant \frac{\sqrt{2}}{3} + \frac{2(1+\log n)}{3\sqrt{n}}.$$

For the off-diagonal entries of V, we know that for each pair $i \neq j$,

$$V_{ij} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{E}\left[d_{ti} d_{tj}\right] = \frac{1}{\sqrt{n}} \mathbb{E}\left[\left(\sum_{t=1}^{n} d_{ti}\right)\left(\sum_{t=1}^{n} d_{tj}\right)\right] = \frac{1}{\sqrt{n}} \mathrm{Cov}\left\{R_{ni}, R_{nj}\right\} \leqslant \frac{\sqrt{2}}{3} + \frac{2(1+\log n)}{3\sqrt{n}},$$

completing the proof. □

Next, we verify the other conditions in Theorem 5.2. Condition (i) of Theorem 5.2 is verified in the following proposition.

Proposition 5.9 (Boundedness of third moments). *The martingale differences $\{\boldsymbol{d}_t : t \in [n]\}$ satisfy the moment bound*

$$\sum_{t=1}^{n} \mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^3\right] \leqslant 4\sqrt{2n} \log n.$$

Proof. Since we always have that $\| \boldsymbol{d}_t \|_\infty \leqslant 1$ (see, e.g., Arlotto et al., 2015, (43)), it suffices for us to work with $\| \boldsymbol{d}_t \|_\infty^2$. From Arlotto et al. (2015, Equation (41), (46)), we know that

$$\mathbb{E}\left[d_{tj}^2 \mid \mathcal{F}_{t-1}\right] \leqslant \int_{X_{(t-1)j}}^{h_{n-t+1}(X_{(t-1)j})} \left\{1 + v_{n-t}(w) - v_{n-t}(X_{(t-1)j})\right\}^2 dw \leqslant h_{n-t+1}(X_{(t-1)j}) - X_{(t-1)j}.$$

As a result,

$$\mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^2\right] = \mathbb{E}\left[\mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^2 \,\middle|\, \mathcal{F}_{t-1}\right]\right] \leqslant \mathbb{E}\left[\max_{j\in[p]} \left(h_{n-t+1}(X_{(t-1)j}) - X_{(t-1)j}\right)\right].$$

Then we apply (5.23) to obtain that

$$\mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^2\right] \leqslant \frac{4\log(n-t+1)+6}{\sqrt{2(n-t+1)}}.$$

Summing up this last inequality yields the higher moment bound

$$\sum_{t=1}^{n} \mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^3\right] \leqslant \sum_{t=1}^{n} \mathbb{E}\left[\| \boldsymbol{d}_t \|_\infty^2\right] \leqslant \sum_{t=1}^{n} \frac{4\log(n-t+1)+6}{\sqrt{2(n-t+1)}} \leqslant 4\sqrt{2n}\log n,$$

so Condition (i) of Theorem 5.2 is satisfied. □

The positive definiteness in Condition (ii) of Theorem 5.2 is verified in the following proposition.

Proposition 5.10 (Positive semidefinite condition). *When the uncertainty distribution F satisfies Assumption 5.1, there exists constant c_1 that depends only on $\bar{c}$ such that, if δ_n satisfies that*

$$\delta_n \geqslant \frac{c_1\theta(n,p)\log^2 p\log^3 n}{n^{1/4-3a/2}},$$

then we have that with probability at least $1-3p^2\exp(-n^{2a})$,

$$V_n \preceq \left(1+\delta_n/\log^2 p\right) V.$$

Proof. We verify that if the distribution F of the uncertainty satisfies Assumption 5.1, then with high probability, the matrix $\left\{\left(1+\delta_n/\log^2 p\right) V - V_n\right\}$ is diagonally dominant, and hence we obtain that

$$V_n \preceq \left(1+\delta_n/\log^2 p\right) V.$$

From Proposition 5.5 we know that with probability at least $1-3p^2\exp(-n^{2a})$ for some fixed $a \in (0, 1/6)$,

$$\| V_n - V \|_\infty \leqslant \frac{c_1\log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}.$$

Since $V_{jj} \geqslant \sqrt{2}/3 - 2/\sqrt{n}$ (cf. Arlotto et al., 2015, Proposition 19), the diagonal entries of the matrix $\left\{\left(1+\delta_n/\log^2 p\right) V - V_n\right\}$ satisfy that with probability at least $1-3p^2\exp(-n^{2a})$,

$$\left\{\left(1+\delta_n/\log^2 p\right) V - V_n\right\}_{jj} \geqslant \frac{\delta_n}{\log^2 p} V_{jj} - \| V_n - V \|_\infty \geqslant \frac{\sqrt{2}\delta_n}{3\log^2 p} - \frac{c_1\log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}.$$

On the other hand, the off-diagonal entries of the matrix $\left\{\left(1+\delta_n/\log^2 p\right)V - V_n\right\}$ has the upper bound that for $i, j \in [p]$ with $i \neq j$,

$$\left|\left(1+\delta_n/\log^2 p\right)V - V_n\right|_{ij} \leqslant \frac{\delta_n}{\log^2 p}\left|V_{ij}\right| + \left\|V_n - V\right\|_\infty \leqslant O\left(\frac{\delta_n \log^3 n}{\log^2 p\sqrt{n}}\right) + \frac{c_1 \log^3 n}{n^{\frac{1}{4}-\frac{3a}{2}}}.$$

Combining the last two inequalities together, we obtain that for all $j \in [p]$,

$$\frac{\left\{\left(1+\delta_n/\log^2 p\right)V - V_n\right\}_{jj}}{\left|\left(1+\delta_n/\log^2 p\right)V - V_n\right|_{ij}} \geqslant \frac{\delta_n n^{\frac{1}{4}-\frac{3a}{2}}}{c_1 \log^2 p \log^3 n} \quad \text{for all } i \in [p], i \neq j.$$

Hence if the uncertainty distribution F satisfies Assumption 5.1, then for each $j \in [p]$, there are at most $\theta(n,p)$ non-zero entries in each row of the matrix $\left\{\left(1+\delta_n/\log^2 p\right)V - V_n\right\}$. The choice of δ_n guarantees that

$$\frac{\delta_n n^{\frac{1}{4}-\frac{3a}{2}}}{c_1 \log^2 p \log^3 n} \geqslant \theta(n,p),$$

so the matrix $\left\{\left(1+\delta_n/\log^2 p\right)V - V_n\right\}$ is diagonally dominant, and hence

$$V_n \preceq \left(1+\delta_n/\log^2 p\right)V.$$

□

Proof of Theorem 5.4

With all the conditions verified, we are ready to prove Theorem 5.4.

Proof. (Proof of Theorem 5.4.) We verify the conditions required in Theorem 5.2. We set the parameters

$$\psi_n = 3p^2 \exp(-n^{2a}) \quad \text{and} \quad \delta_n = \frac{\log^{3/2} n \log(np)}{n^{1/8-3a/4}}.$$

Then with the parameters satisfying (5.24), one can easily check that: Condition (i) is verified by Proposition 5.9; Condition (ii) is verified by Proposition 5.5, Proposition 5.8, and Proposition 5.10.

The first half of Condition (iii) is verified by Proposition 5.5. Since we can directly compute the martingale differences $\boldsymbol{d}_t$ with observations, in the second half of Condition (iii) in Theorem 5.2, the estimator $\widehat{Z}_{kj}$ is the same as the actual Z_{kj}. As a result, the second half of Condition (iii) is automatically satisfied. □

5.7 Concluding Remarks

In this chapter, we study a general framework of finite-horizon Markov decision problems, and focus on the specific statistical inference problems associated with the optimal and near-optimal policies of such Markov decision problems. The two inferences we focus on are simultaneous hypothesis testing and constructing simultaneous confidence intervals for the total rewards collected by each decision maker. Such inference problems are studied in the high-dimensional regime in which the traditional central limit theorems fail. In particular, we study two different types of MDPs: one with regenerative states and the other with absorbing states. Suitable conditions are given to guarantee asymptotic validity of bootstrap based simultaneous hypothesis testing and simultaneous confidence interval constructions for such two types of MDPs. The key insight from the two applications is that, when one has a comprehensive understanding about the Markov decision problems at hand, it is possible to monitor the implementations of policies of many decision makers simultaneously without knowledge of the correlation structure of the uncertainty they face.

Chapter 6

Conclusion

In this dissertation, we have studied several sequential decision problems under uncertainty. The focus is the total reward collected by different sequential policies. We have analyzed such object from different perspectives. Besides the standard perspective–the expected total reward, we have also analyzed the variance and the limiting distribution of the total reward collected by different sequential policies for certain problems. In addition, we use bootstrap procedure to construct simultaneous confidence intervals for the total rewards collected by each decision maker of many sequential decision problems that are run in parallel. The analyses have both theoretical and practical insights for decision makers facing related problems. Specifically, from a theoretical perspective, our analyses calls for careful comparison when it comes to distinguishing different policies among the class of asymptotically optimal policies for certain problems. The analyses also indicate different risks involved among all asymptotically optimal policies. From a practical perspective, our analyses can be applied to the setting of many sequential decision problems that are correlated and are faced in parallel by self-interested decision makers. The proposed bootstrap procedure can be used for simultaneously testing whether each decision maker is implementing designated policy through the observations of the rewards they collect over time. The tools and techniques developed in this dissertation could potentially find different applications in other related problems.

Bibliography

Andrews, Donald W. K. 2004. The block–block bootstrap: improved asymptotic refinements. *Econometrica* 72(3) 673–700.

Arlotto, Alessandro, Itai Gurvich. 2019. Uniformly bounded regret in the multi-secretary problem. *Stochastic Systems* 9(3) 231–260.

Arlotto, Alessandro, Elchanan Mossel, J. Michael Steele. 2016. Quickest online selection of an increasing subsequence of specified size. *Random Struct. Algor.* 49(2) 235–252.

Arlotto, Alessandro, Vinh V. Nguyen, J. Michael Steele. 2015. Optimal online selection of a monotone subsequence: a central limit theorem. *Stochastic Process. Appl.* 125(9) 3596–3622.

Arlotto, Alessandro, J. Michael Steele. 2011. Optimal sequential selection of a unimodal subsequence of a random sequence. *Combin. Probab. Comput.* 20(6) 799–814.

Arlotto, Alessandro, J. Michael Steele. 2016. A central limit theorem for temporally nonhomogenous Markov chains with applications to dynamic programming. *Math. Oper. Res.* 41(4) 1448–1468.

Arlotto, Alessandro, Yehua Wei, Xinchang Xie. 2018. An adaptive $O(\log n)$-optimal policy for the online selection of a monotone subsequence from a random sample. *Random Struct. Algor.* 52(1) 41–53.

Arlotto, Alessandro, Xinchang Xie. 2020a. Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards. *Stochastic Systems* forthcoming.

Arlotto, Alessandro, Xinchang Xie. 2020b. Sequential policies and the distributions of their total rewards in dynamic and stochastic knapsack problems. *Working paper – Duke University* .

Balseiro, Santiago R., David B. Brown. 2019. Approximations to stochastic dynamic programs via information relaxation duality. *Oper. Res.* 67(2) 577–597.

Baruah, Sanjoy, Jayant Haritsa, Nitin Sharma. 1994. On-line scheduling to maximize task completions. *1994 Proceedings Real-Time Systems Symposium*. IEEE, San Juan, PR, 228–236.

Baryshnikov, Yuliy M., Alexander V. Gnedin. 2000. Sequential selection of an increasing sequence from a multidimensional random sample. *Ann. Appl. Probab.* 10(1) 258–267.

Bellman, Richard. 1957. *Dynamic programming*. Princeton University Press, Princeton, N. J.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, Christian Bailey Hansen, Kengo Kato. 2018. High-dimensional econometrics and regularized GMM. *ArXiv e-print 1806.01888* .

Belloni, Alexandre, Roberto I. Oliveira. 2017. Approximate group context tree. *Ann. Statist.* 45(1) 355–385.

Belloni, Alexandre, Roberto I. Oliveira. 2018. A high dimensional central limit theorem for martingales, with applications to context tree models. *ArXiv e-print 1809.02741* .

Belloni, Alexandre, Xinchang Xie. 2020. Data-driven monitoring the optimality of policies across many markov decision problems. *Working paper – Duke University* .

Bernstein, Serge. 1927. Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* 97(1) 1–59.

Bertsekas, Dimitri P. 2005. *Dynamic Programming and Optimal Control*, vol. I,II. Athena Scientific, Belmont, MA.

Bhalgat, Anand, Ashish Goel, Sanjeev Khanna. 2011. Improved approximation results for stochastic knapsack problems. *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA, 1647–1665.

Blado, Daniel, Weihong Hu, Alejandro Toriello. 2016. Semi-infinite relaxations for the dynamic knapsack problem with stochastic item sizes. *SIAM J. Optim.* 26(3) 1625–1648.

Blado, Daniel, Alejandro Toriello. 2019. Relaxation analysis for the dynamic knapsack problem with stochastic item sizes. *SIAM J. Optim.* 29(1) 1–30.

Boshuizen, Frans A, Robert P Kertz. 1999. Smallest-fit selection of random sizes under a sum constraint: weak convergence and moment comparisons. *Adv. in Appl. Probab.* 31(1) 178–198.

Bradley, Richard C. 2005. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2 107–144.

Bruss, F. Thomas, Freddy Delbaen. 2001. Optimal rules for the sequential selection of monotone subsequences of maximum expected length. *Stochastic Process. Appl.* 96(2) 313–342.

Bruss, F. Thomas, Freddy Delbaen. 2004. A central limit theorem for the optimal selection process for monotone subsequences of maximum expected length. *Stochastic Process. Appl.* 114(2) 287–311.

Bruss, F. Thomas, Mitia Duerinckx. 2015. Resource dependent branching processes and the envelope of societies. *Ann. Appl. Probab.* 25(1) 324–372.

Bruss, F. Thomas, James B. Robertson. 1991. "Wald's lemma" for sums of order statistics of i.i.d. random variables. *Adv. in Appl. Probab.* 23(3) 612–623.

Bumpensanti, Pornpawee, He Wang. 2020. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Sci.* forthcoming.

Carraway, Robert L., Robert L. Schmidt, Lawrence R. Weatherford. 1993. An algorithm for maximizing target achievement in the stochastic knapsack problem with normal returns. *Naval Res. Logist.* 40(2) 161–173.

Cayley, Arthur. 1875. Mathematical questions and their solutions. *Educational Times* 22 18–19, see *The Collected Mathematical Papers of Arthur Cayley*, 10, 587–588, (1986). Cambridge University Press, Cambridge.

Chen, Lijian, Tito Homem-de Mello. 2010. Re-solving stochastic programming models for airline revenue management. *Ann. Oper. Res.* 177(1) 91–114.

Chen, Xiaohong, Qi-Man Shao, Wei Biao Wu, Lihu Xu. 2016. Self-normalized Cramér-type moderate deviations under dependence. *Ann. Statist.* 44(4) 1593–1617.

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* 41(6) 2786–2819.

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato. 2015. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* 162(1-2) 47–70.

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato. 2017. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* 45(4) 2309–2352.

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato. 2019. Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* 86(5) 1867–1900.

Coffman, E. G., Jr., L. Flatto, R. R. Weber. 1987. Optimal selection of stochastic intervals under a sum constraint. *Adv. in Appl. Probab.* 19(2) 454–473.

Cooper, William L. 2002. Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50(4) 720–727.

Dantzig, George B. 1957. Discrete-variable extremum problems. *Oper. Res.* 5(2) 266–277.

Dean, Brian C., Michel X. Goemans, Jan Vondrák. 2004. Approximating the stochastic knapsack problem: the benefit of adaptivity. *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Press, Piscataway, NJ, 208–217. doi:10.1109/FOCS.2004.15.

Dean, Brian C., Michel X. Goemans, Jan Vondrák. 2005. Adaptivity and approximation for stochastic packing problems. *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, NY, 395–404.

Dean, Brian C., Michel X. Goemans, Jan Vondrák. 2008. Approximating the stochastic knapsack problem: the benefit of adaptivity. *Math. Oper. Res.* 33(4) 945–964.

Derman, C., G. J. Lieberman, S. M. Ross. 1975. A stochastic sequential allocation model. *Oper. Res.* 23(6) 1120–1130.

Derman, C., G. J. Lieberman, S. M. Ross. 1978. A renewal decision problem. *Management Sci.* 24(5) 554–561.

Dobrushin, Roland L. 1956a. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications* 1(1) 65–80.

Dobrushin, Roland L'vovich. 1956b. Central limit theorem for nonstationary Markov chains. II. *Theory of Probability & Its Applications* 1(4) 329–383.

Fan, Jianqing, Qiwei Yao. 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Fryzlewicz, Piotr, Suhasini Subba Rao. 2011. Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* 17(1) 320–346.

Gallego, Guillermo, Garrett van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8) 999–1020.

Gallego, Guillermo, Garrett van Ryzin. 1997. A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.* 45(1) 24–41.

Gnedin, Alexander, Amirlan Seksenbayev. 2019. Asymptotics and renewal approximation in the online selection of increasing subsequence. *ArXiv e-print 1904.11213* .

Gnedin, Alexander V. 1999. Sequential selection of an increasing subsequence from a sample of random size. *J. Appl. Probab.* 36(4) 1074–1085.

Gupta, Anupam, Ravishankar Krishnaswamy, Marco Molinaro, R. Ravi. 2011. Approximation algorithms for correlated knapsacks and non-martingale bandits. *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*. IEEE Computer Soc., Los Alamitos, CA, 827–836.

Henig, Mordechai I. 1990. Risk criteria in a stochastic knapsack problem. *Oper. Res.* 38(5) 820–825.

Iglehart, Donald L. 1963. Optimality of (s, S) policies in the infinite horizon dynamic inventory problem. *Management Sci.* 9(2) 259–267.

Ilhan, Taylan, Seyed M. R. Iravani, Mark S. Daskin. 2011. TECHNICAL NOTE—The adaptive knapsack problem with stochastic rewards. *Oper. Res.* 59(1) 242–248.

Jarner, Søren F., Gareth O. Roberts. 2002. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* 12(1) 224–247.

Jasin, Stefanus, Sunil Kumar. 2012. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Math. Oper. Res.* 37(2) 313–345.

Jasin, Stefanus, Sunil Kumar. 2013. Analysis of deterministic LP-based booking limit and bid price controls for revenue management. *Oper. Res.* 61(6) 1312–1320.

Jones, Galin L. 2004. On the Markov chain central limit theorem. *Probability Surveys* 1 299–320.

Kellerer, Hans, Ulrich Pferschy, David Pisinger. 2004. *Knapsack problems*. Springer-Verlag, Berlin.

Kleinberg, Robert. 2005. A multiple-choice secretary algorithm with applications to online auctions. *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, NY, 630–631.

Kleywegt, Anton J., Jason D. Papastavrou. 1998. The dynamic and stochastic knapsack problem. *Oper. Res.* 46(1) 17–35.

Kleywegt, Anton J., Jason D. Papastavrou. 2001. The dynamic and stochastic knapsack problem with random sized items. *Oper. Res.* 49(1) 26–41.

Lahiri, Soumendra Nath. 2003. *Resampling methods for dependent data*. Springer Series in Statistics, Springer-Verlag, New York.

Li, Jian, Wen Yuan. 2013. Stochastic combinatorial optimization via Poisson approximation. *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing*. ACM, New York, NY, 971–980.

Logan, B. F., L. A. Shepp. 1977. A variational problem for random Young tableaux. *Advances in Math.* 26(2) 206–222.

Lueker, George S. 1998. Average-case analysis of off-line and on-line knapsack problems. *J. Algorithms* 29(2) 277–305.

Ma, Will. 2018. Improvements and generalizations of stochastic knapsack and Markovian bandits approximation algorithms. *Math. Oper. Res.* 43(3) 789–812.

Mandl, Petr, Monika Laušmanová. 1991. Two extensions of asymptotic methods in controlled Markov chains. *Annals of Operations Research* 28(1) 67–79.

Marchetti-Spaccamela, A., C. Vercellis. 1995. Stochastic on-line knapsack problems. *Math. Program.* 68(1, Ser. A) 73–104.

Martello, Silvano, Paolo Toth. 1990. *Knapsack problems*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Ltd., Chichester.

Maxwell, Michael, Michael Woodroofe. 2000. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* 28(2) 713–724.

McLeish, D. L. 1974. Dependent central limit theorems and invariance principles. *Ann. Probab.* 2(4) 620–628.

Mendoza-Pérez, Armando F. 2008. Asymptotic normality of average cost Markov control processes. *Morfismos* 12(2) 33–52.

Mendoza-Pérez, Armando F., Onésimo Hernández-Lerma. 2010. Asymptotic normality of discrete-time Markov control processes. *J. Appl. Probab.* 47(3) 778–795.

Merzifonluoğlu, Yasemin, Joseph Geunes, H. Edwin Romeijn. 2012. The static stochastic knapsack problem with normally distributed item sizes. *Math. Program.* 134(2, Ser. A) 459–489.

Meyn, Sean, Richard L. Tweedie. 2009. *Markov Chains and Stochastic Stability*. 2nd ed. Cambridge University Press, Cambridge.

Moser, Leo. 1956. On a problem of Cayley. *Scripta Mathematica* 22 289–292.

Nagaev, S.V. 1976. An estimate of the remainder term in the multidimensional central limit theorem. *Proceedings of the Third Japan—USSR Symposium on Probability Theory*. Springer, 419–438.

Nakai, Toru. 1986. An optimal selection problem for a sequence with a random number of applicants per period. *Oper. Res.* 34(3) 478–485.

Paparoditis, Efstathios, Dimitris N. Politis. 2001. Tapered block bootstrap. *Biometrika* 88(4) 1105–1119.

Paparoditis, Efstathios, Dimitris N. Politis. 2002. The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.* 5(1) 131–148.

Papastavrou, Jason D., Srikanth Rajagopalan, Anton J. Kleywegt. 1996. The dynamic and stochastic knapsack problem with deadlines. *Management Sci.* 42(12) 1706–1718.

Peligrad, Magda. 1985. An invariance principle for φ-mixing sequences. *Ann. Probab.* 13(4) 1304–1313.

Peligrad, Magda. 1990. On Ibragimov–Iosifescu conjecture for φ-mixing sequences. *Stochastic Process. Appl.* 35(2) 293–308.

Peligrad, Magda. 2012. Central limit theorem for triangular arrays of non-homogeneous Markov chains. *Probability Theory and Related Fields* 154(3-4) 409–428.

Peng, Peichao, J. Michael Steele. 2016. Sequential selection of a monotone subsequence from a random permutation. *Proc. Amer. Math. Soc.* 144(11) 4973–4982.

Powell, Warren B. 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. 2nd ed. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

Prastacos, Gregory P. 1983. Optimal sequential investment decisions under conditions of uncertainty. *Management Sci.* 29(1) 118–134.

Puterman, Martin L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication.

Reiman, Martin I., Qiong Wang. 2008. An asymptotically optimal policy for a quantity-based network revenue management problem. *Math. Oper. Res.* 33(2) 257–282.

Rhee, WanSoo, Michel Talagrand. 1991. A note on the selection of random variables under a sum constraint. *J. Appl. Probab.* 28(4) 919–923.

Samuels, Stephen M., J. Michael Steele. 1981. Optimal sequential selection of a monotone sequence from a random sample. *Ann. Probab.* 9(6) 937–947.

Scarf, Herbert. 1959. The optimality of (S, s) policies in the dynamic inventory problem. Kenneth Joseph Arrow, Patrick Suppes, Samuel Karlin, eds., *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, California, 196–202.

Seksenbayev, Amirlan. 2018. Refined asymptotics in the online selection of an increasing subsequence. *ArXiv e-print 1808.06300* .

Sklar, M. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8 229–231.

Steele, J. Michael. 2016. The Bruss-Robertson inequality: elaborations, extensions, and applications. *Math. Appl. (Warsaw)* 44(1) 3–16.

Talluri, Kalyan, Garrett van Ryzin. 1998. An analysis of bid-price controls for network revenue management. *Management Sci.* 44(11-part-1) 1577–1593.

Talluri, Kalyan T., Garrett J. van Ryzin. 2004. *The theory and practice of revenue management*. International Series in Operations Research & Management Science, 68, Kluwer Academic Publishers, Boston, MA.

Utev, Sergei Aleksandrovich. 1991. On the central limit theorem for φ-mixing arrays of random variables. *Theory of Probability & Its Applications* 35(1) 131–139.

Vera, Alberto, Siddhartha Banerjee. 2019. The Bayesian prophet: A low-regret framework for online decision making. *SIGMETRICS Perform. Eval. Rev.* 47(1) 81–82.

Veršik, A. M., S. V. Kerov. 1977. Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Dokl. Akad. Nauk SSSR* 233(6) 1024–1027.

White, Douglas J. 1993. A survey of applications of Markov decision processes. *Journal of the Operational Research Society* 44(11) 1073–1096.

Wu, Huasen, R. Srikant, Xin Liu, Chong Jiang. 2015. Algorithms with logarithmic or sublinear regret for constrained contextual bandits. C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett, eds., *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 433–441.

Wu, Wei Biao. 2005. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* 102(40) 14150–14154.

Yu, Bin. 1994. Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* 22(1) 94–116.

Zhang, Danna, Wei Biao Wu. 2017. Gaussian approximation for high dimensional time series. *Ann. Statist.* 45(5) 1895–1919.

Zhang, Xianyang, Guang Cheng. 2014. Bootstrapping high dimensional time series. *ArXiv e-print 1406.1037* .

Zhang, Xianyang, Guang Cheng. 2018. Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* 24(4A) 2640–2675.