

Platform Operations: From Models to Methods

Akshit Kumar

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
under the Executive Committee  
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

© 2025

Akshit Kumar

All Rights Reserved

# Abstract

Platform Operations: From Models to Methods

Akshit Kumar

Online platforms – from on-demand home-services and e-commerce fulfillment networks to content streamers – thrive on their ability to *match* demand with supply. Doing so well requires algorithms that cope with uncertainty, high-dimensional type spaces, mis-aligned and multiple objectives, and informational frictions. This dissertation develops models and methods that illuminate these challenges and propose simple, and often near-optimal solutions. The dissertation is organized into five self-contained but thematically linked chapters – the first three chapters are related to problems in online matching, dynamic resource allocation and multi-objective optimization in order fulfillment problems, while, the last two chapters deal with design and optimization of recommendation systems.

First, we tackle algorithm design questions that arise in online matching and dynamic resource allocation problems. How should a centralized matching platform dynamically match heterogeneous service providers with heterogeneous customers? What are the fundamental drivers of algorithmic performance in these dynamic resource allocation platforms? Is there a unifying algorithmic principle which can solve large class of dynamic resource allocation problems? How to make resource allocation decisions with multiple and often competing objectives? In Chapter 1, we study dynamic two-sided matching with heterogeneous demand and supply modeled as highly dimensional weight and feature vectors respectively. We show that simple myopic policies such as Greedy which are practically prevalent can impact be highly sub-optimal. We develop a forward-looking supply aware policy dubbed Simulate-Optimize-Assign-Repeat (SOAR). We prove that SOAR achieves the optimal regret scaling under different assumptions on the demand and supply distributions. In Chapter 2, we broaden the scope to general online resource allocation problems. We identify a novel driver of algorithmic performance – the spatial distribution of demand types. We develop a unifying algorithm dubbed Repeatedly Act using Multiple Simulations (RAMS)

which is a generalization of SOAR studied in Chapter 1. In Chapter 3, we turn to multi-objective optimization in the context of order fulfillment problems. We develop a principled framework for weight generation to enable the weighted objective approach.

Next, we take the viewpoint of a system designer and tackle platform design questions in the context of recommendation systems. By optimizing for measurable proxies, are recommendation systems at risk of significantly under-delivering on utility? If so, how can one improve utility which is seldom measured? Different information provisioning tools exist, such as public rankings and personalized recommendations, but when do these tools work and when do they not? What is the role of the market setting in driving the efficacy of these different information provisioning tools? In Chapter 4, we study a stylized model of repeated user consumption. We demonstrate that optimizing for measurable proxies like engagement can lead to significant utility losses. Instead, we propose a utility-aware policy that initially recommends diverse set of options. As the platform becomes more forward-looking, our utility-aware policy achieves the best of both worlds: near-optimal utility and near-optimal engagement simultaneously. Our study elucidates an important feature of recommendation systems; given the ability to suggest multiple items, one can perform significant exploration without incurring significant reductions in engagement. By recommending high-risk, high-reward items alongside popular items, systems can enhance discovery of high utility items without significantly affecting engagement. In Chapter 5, we ask when personalized recommendations are worth their added complexity relative to public rankings. In unconstrained supply settings, both public rankings and personalized recommendations improve welfare, with their relative value determined by the degree of preference heterogeneity. In contrast, in supply-constrained settings, revealing just the common term of the utility, as done by public rankings, provides limited benefit since the total common value available is limited by capacity constraints, whereas personalized recommendations, by revealing both common and idiosyncratic terms, significantly enhance welfare by enabling agents to match with items they idiosyncratically value highly.

## Table of Contents

Acknowledgments . . . . .	xi
Dedication . . . . .	xiii
Executive Summary . . . . .	1
0.1 Online matching and Dynamic Resource Allocation: Models and Algorithms . . .	2
0.1.1 Chapter 1: Feature-Based Dynamic Matching . . . . .	2
0.1.2 Chapter 2: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances . . . . .	4
0.1.3 Chapter 3: MOTIF: Multi-Objective Tradeoff In Fulfillment . . . . .	5
0.2 Design of Recommendation Systems: Models and Insights . . . . .	6
0.2.1 Chapter 4: On the Perils of Optimizing the Measurable . . . . .	6
0.2.2 Chapter 5: Impact of Rankings and Personalized Recommendations in Marketplaces . . . . .	8
Chapter 1: Feature-Based Dynamic Matching . . . . .	10
1.1 Introduction . . . . .	10
1.1.1 Main Contribution . . . . .	12
1.1.2 Related Literature . . . . .	15
1.1.3 Notation . . . . .	18
1.2 Model . . . . .	19

1.2.1	Modelling scarce supply and rejection cost . . . . .	21
1.3	SOAR: Algorithmic Principle and Performance Analysis . . . . .	21
1.3.1	Insufficiency of Myopic Policies . . . . .	22
1.3.2	Simulate, Optimize, Assign, Repeat (SOAR) Principle . . . . .	23
1.4	Near-optimal Regret Scaling of SOAR for the $-\ X - Y\ ^p$ and $\langle X, Y \rangle$ Quality Functions . . . . .	29
1.4.1	Performance Guarantees for $-\ X - Y\ ^p$ Quality Functions . . . . .	30
1.4.2	Performance Guarantees for the $\langle X, Y \rangle$ Quality Function . . . . .	32
1.5	Numerics . . . . .	35
1.5.1	Setting (I) $P = Q = \text{Uniform}([0, 1]^d)$ . . . . .	36
1.5.2	Setting (II) $P = \text{Uniform}([0, 1]^d), Q = \text{Uniform}([0, 2]^d)$ . . . . .	36
1.5.3	Setting (III) $P = \text{TruncNorm}(\mu, \Sigma), Q = \text{Uniform}([0, 1]^d)$ . . . . .	38
1.5.4	Setting (IV) $P = \text{Uniform}([0, 1]^d), Q = \text{TruncNorm}(\mu_{d-2 \times 1}, \Sigma_{d-2 \times d-2}) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$ . . . . .	39
1.6	Conclusion and Future Research . . . . .	39
Chapter 2: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances . . . . .		
2.1	Introduction . . . . .	41
2.1.1	Related Literature . . . . .	44
2.2	Model . . . . .	46
2.3	Fundamental Limits on Achievable Performance . . . . .	49
2.3.1	General Class of Distributions For the multisecretary Problem . . . . .	50
2.3.2	Fundamental Lower bound on Performance . . . . .	53
2.4	Algorithmic Design Principles for Near Optimal Performance . . . . .	55

2.4.1	Failure of the CE policy under many types with gaps . . . . .	55
2.4.2	Conservativeness with respect to gaps . . . . .	56
2.4.3	Achieving Conservativeness with respect to Gaps via a Simulation-based Policy . . . . .	61
2.5	Unifying Algorithm: Repeatedly Act using Multiple Simulations . . . . .	63
2.5.1	Algorithmic Description . . . . .	64
2.5.2	Performance Analysis: Meta Theorem for RAMS . . . . .	66
2.5.3	Connection of RAMS to prior work . . . . .	68
2.5.4	Numerical Simulations . . . . .	69
2.6	Conclusion . . . . .	72
Chapter 3: MOTIF: Multi-Objective Tradeoff in Fulfillment . . . . .		74
3.1	Introduction . . . . .	74
3.2	Pitfalls of Cost Relaxation (CR) . . . . .	77
3.2.1	Inconsistent Objective Tradeoff in CR leads to Pareto Inefficiency . . . . .	77
3.2.2	Increased Complexity and Computational Overheads . . . . .	78
3.3	Blended Objective (BO) Approach: Theoretical Foundations . . . . .	79
3.3.1	Notation and Definition . . . . .	79
3.3.2	Blended Objective (BO) Formulation . . . . .	80
3.4	Principled Framework for obtaining CR-dominating solutions . . . . .	81
3.5	Pareto Frontier: A decision support tool for business leaders . . . . .	82
3.6	Wide-scale Rollout and Performance . . . . .	84
Chapter 4: The Fault in Our Recommendations: On the Perils of Optimizing the Measurable . . . . .		85

4.1	Introduction . . . . .	85
4.1.1	Related Literature . . . . .	89
4.2	Model . . . . .	90
4.3	Analysis of the Two-point distribution for the niche type . . . . .	92
4.4	Robustness of Insights Under General Settings . . . . .	96
4.5	Conclusion . . . . .	99
Chapter 5: Impact of Rankings and Personalized Recommendations in Marketplaces . . . .		102
5.1	Introduction . . . . .	102
5.1.1	Main Contributions . . . . .	106
5.1.2	Related Literature . . . . .	110
5.2	Model . . . . .	112
5.3	Main Results . . . . .	114
5.3.1	Uncapacitated supply setting . . . . .	115
5.3.2	Capacitated supply setting . . . . .	122
5.4	Proof of Theorems for utility distributions with Pareto tail . . . . .	125
5.4.1	Useful Results . . . . .	126
5.4.2	Uncapacitated Supply Setting . . . . .	126
5.4.3	Capacitated supply setting . . . . .	129
5.5	Conclusion . . . . .	132
References . . . . .		134
Appendix A: Feature-Based Dynamic Matching . . . . .		146

A.1	Proof of $U_\infty(P, Q, \varphi) \geq U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$ . . . . .	146
A.2	Proof of the Failure of Greedy in Section 1.3.1 . . . . .	147
A.2.1	Proof of Proposition 1 . . . . .	147
A.3	Proof of Corollary 1 . . . . .	152
A.4	Proof of Corollary 2 . . . . .	153
A.5	Examples of Matching Instances Scale Regularly . . . . .	155
A.6	Details Related to Optimal Transport and Useful Known Results . . . . .	156
A.6.1	Background on Optimal Transport and Wasserstein- $p$ distance . . . . .	156
A.6.2	Existing Results on convergence of Empirical Optimal Transport value . . . . .	159
A.6.3	Equivalence of $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$ and $\varphi_2(X, Y) = -\ X - Y\ ^2$ . . . . .	161
A.7	Proof of Theorem 2 . . . . .	163
A.8	Proof of Proposition 2 . . . . .	165
A.9	Vanishing Regret for polynomial kernel quality function . . . . .	169
A.10	Proof of Theorem 3 . . . . .	170
Appendix B: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances . . . . . 172		
B.1	Proof of Theorem 4 . . . . .	172
B.1.1	Proof of Lemma 11 . . . . .	178
B.1.2	Proof of Lemma 12 . . . . .	180
B.1.3	Proof of Lemma 13 . . . . .	182
B.1.4	Proof of Lemma 14 . . . . .	183
B.2	Details and Analysis of CWG Policy . . . . .	184
B.2.1	Phase Structure of Algorithm 2 . . . . .	184

B.2.2	Hindsight To Go (HTG) and HTG Threshold . . . . .	185
B.2.3	Proof Outline . . . . .	186
B.2.4	Preliminaries and Helper Lemmas . . . . .	187
B.2.5	Formal Proof of Theorem 5 . . . . .	188
B.2.6	Proof of Helper Lemmas . . . . .	194
B.3	Proof of Corollary 5 . . . . .	195
B.4	Proof of Corollary 6 . . . . .	196
B.5	Recovering existing regret guarantees for RAMS . . . . .	197
B.6	Proofs Related to RAMS . . . . .	200
B.6.1	Proof of Claim 1 . . . . .	200
B.6.2	Proof of Lemma 2 . . . . .	200
B.6.3	Proof of Theorem 6 . . . . .	200
B.6.4	Proof of Corollaries 12, 13 and 14 . . . . .	204
B.7	Relating the order fulfillment problem to the multisecretary problem . . . . .	206
B.7.1	Motivating Example . . . . .	206
B.7.2	Stylized model of order fulfillment . . . . .	207
B.8	Approximation of a distribution by $(\beta, \varepsilon_0, \delta)$ -clustered distributions . . . . .	210
B.9	Representation through $(\beta, \varepsilon_0, \delta)$ -clustered distributions . . . . .	210
Appendix C:	The Fault in Our Recommendations: On the Perils of Optimizing the Mea- surable . . . . .	214
C.1	Useful Technical Result . . . . .	214
C.2	Proof of Theorem 7 . . . . .	214
C.3	Proof of Theorem 8 . . . . .	215

C.4	Proof of Helper Lemmas . . . . .	217
C.4.1	Proof of Lemma 20 . . . . .	217
C.4.2	Proof of Lemma 21 . . . . .	218
Appendix D: Impact of Rankings and Personalized Recommendations in Marketplaces . .		221
D.1	Proof of intermediate results . . . . .	221
D.1.1	Proof of Proposition 4 . . . . .	221
D.1.2	Proof of Lemma 3 . . . . .	223
D.1.3	Proof of Lemma 4 . . . . .	225
D.2	Proof of Theorems for utility distributions with Exponential tail . . . . .	227
D.2.1	Connection between Pareto and Exponential tail . . . . .	227
D.2.2	Useful Intermediate Results . . . . .	227
D.2.3	Proof of Theorem 10 . . . . .	232
D.2.4	Proof of Theorem 12 . . . . .	233
D.3	(Partial) Results for bounded utility distributions . . . . .	234

## List of Figures

1.1	Comparing the performance of Hierarchical Greedy (HG), Greedy and SOAR for $P = Q = \text{Uniform}([0, 1]^d)$ . . . . .	37
1.2	Comparing the performance of Greedy, OT + Greedy and SOAR for $P = \text{Uniform}([0, 1/2]^d)$ and $Q = \text{Uniform}([0, 1]^d)$ . . . . .	38
1.3	Comparing the performance of Greedy and SOAR for $P = \text{TruncNorm}(\mu, \Sigma)$ and $Q = \text{Uniform}([0, 1]^d)$ . . . . .	38
1.4	Comparing the performance of Greedy and SOAR for $P = \text{Uniform}([0, 1]^d)$ and $Q = \text{TruncNorm}(\mu, \Sigma) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$ . . . . .	39
2.1	The PDF and CDF for the bimodal distributions $F_\beta$ . Notice the gap from 1/4 to 3/4. . . . .	52
2.2	Implementation of the CWG principle using two algorithmic approaches. . . . .	62
2.3	(a) Illustrates the performance of CWG for different distributions, (b) compares the performance of the CE, CWG and RAMS policies on $F_0 = \text{Unif}([0, 1/4] \cup [3/4, 1])$ , (c) highlights the polynomial regret scaling (with the exponent dependent on $\beta$ ) for the gapless variant of the $F_\beta$ distribution given in (2.1), (d) compares the performance of IRT, Bayes Selector and RAMS for NRM with a few types. . . . .	71
4.1	A comparison of different policies showing that it is possible to improve utility substantially from an engagement-maximizing policy with minimal loss in engagement. Note that the discount factor is $\delta = 0.99$ . . . . .	88
4.2	CDF of Generalized Pareto Distribution for $\mu = -1$ and $\sigma = 1 - \xi$ . . . . .	97
4.3	Impact of $\xi$ on utility and engagement for $\delta = 0$ . . . . .	99
4.4	Impact of $\xi$ on utility and engagement for $\delta = 0.999$ . . . . .	100

5.1	Different information regimes studied in this work . . . . .	106
5.2	Shows the marginal impact of public rankings and personalized recommendations and their interplay with (i) capacity constraints (in the rows) and (ii) level of heterogeneity (in the columns). Low level of heterogeneity refers $\rho \in (0, 1/2)$ and high level of heterogeneity refers to $\rho \in (1/2, 1)$ . . . . .	109
5.3	(a) Simulation plot of $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)/c_q \Gamma(1 - \alpha_q) \cdot n^{1/\alpha_q}$ as a function of $\rho \in [0, 1]$ where $P_q$ and $P_\varphi$ are Pareto distributions with $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 2$ , (b) Plot of $g(\rho; \alpha)$ for different values of $\alpha$ , (c) Simulation plot of $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/(\Gamma(1 - 1/\alpha)n^{1/\alpha})$ as a function of $\rho \in [0, 1]$ where $P_q$ and $P_\varphi$ are Pareto distributions with $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 2$ , (d) Simulation plot of $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/(\Gamma(1 - 1/\alpha)n^{1/\alpha})$ as a function of $\rho \in [0, 1]$ where $P_q$ and $P_\varphi$ are Pareto distributions with $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 5$ . . . . .	119
5.4	Simulation plot of (a) $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)/\ln n$ and (b) $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/\ln n$ as a function of $\rho \in [0, 1]$ where $P_q$ and $P_\varphi$ are exponential distributions with rate $\lambda_q = \lambda_\varphi = 1$ . . . . .	121
5.5	Simulation plot of $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$ and $\Delta_{q \rightarrow u}^{\text{cap}}(n)/C_\varphi n^{1/\alpha_\varphi}$ as a function of $\rho \in [0, 1]$ when $P_q$ and $P_\varphi$ are the Pareto distribution with exponent $\alpha_q = \alpha_\varphi = 2$ and $c_q = c_\varphi = 1$ . . . . .	124
5.6	Simulation plot of $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$ and $\Delta_{q \rightarrow u}^{\text{cap}}(n)/\ln n$ as a function of $\rho \in [0, 1]$ when $P_q$ and $P_\varphi$ are the exponential distribution with rate $\lambda_q = \lambda_\varphi = 1$ . . . . .	125
B.1	Partition of the set $\mathcal{S} = [0, \ell] \cup [u, 1]$ into disjoint set $\mathcal{I}_L = [0, \ell_2), \mathcal{I}_{M_1} = [\ell_2, \ell_1), \mathcal{I}_{M_2} = [\ell_1, \ell], \mathcal{I}_{M_3} = [u, u_1), \mathcal{I}_{M_4} = [u_1, u_2), \mathcal{I}_H = [u_2, 1]$ , where $\ell_1 \triangleq \ell - \Delta_\beta, \ell_2 \triangleq \ell - c_0 \Delta_\beta, u_1 \triangleq u + \Delta_\beta, u_2 \triangleq u + c_0 \Delta_\beta$ and $\tilde{\Delta}_\beta \triangleq (c_0 - 1) \Delta_\beta$ . . . . .	173
B.2	Illustration of spatially distributed demand with two fulfillment centers for the order fulfillment problem . . . . .	207
B.3	Stylized example for order fulfillment with no demand from region $\mathfrak{R}_2$ . . . . .	208
B.4	(Left) PDF $f_\beta$ of distribution $F$ (Center) a few types (Right) many small types . . . . .	212
D.1	Simulations plot of $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$ and $\Delta_{q \rightarrow u}^{\text{uncap}}(n)$ as a function of $\rho \in [0, 1]$ when $P_q$ and $P_\varphi$ are the Uniform( $[0, 1]$ ). . . . .	235
D.2	Simulations plot of $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$ and $\Delta_{q \rightarrow u}^{\text{cap}}(n)$ as a function of $\rho \in [0, 1]$ when $P_q$ and $P_\varphi$ are the Uniform( $[0, 1]$ ). . . . .	237

## List of Tables

1.1	The average regret scaling for $\varphi(X, Y) = \langle X, Y \rangle$ . Here $a \wedge b := \min\{a, b\}$ . . . . .	15
1.2	Summary of algorithms considered for numerical simulations . . . . .	36
3.1	Demand and Fulfillment Options . . . . .	77
4.1	Loss in engagement and gain in utility of PEAR compared to the engagement-maximizing policy APP . . . . .	88

## Acknowledgements

These five years at Columbia Business School have been the most formative of my life, and I feel profoundly grateful to the people who guided, challenged, and supported me throughout the journey.

First and foremost, I owe an immeasurable debt of gratitude to my advisors, Omar Besbes and Yash Kanoria. Omar's encyclopedic command of the literature, his instinct for asking the big, important questions, and his gift for connecting seemingly disparate ideas have continually broadened my intellectual horizons. Yash's razor-sharp intuition – and his uncanny ability to generate quick, almost invariably correct conjectures – never cease to amaze me. Beyond their scholarly brilliance, both are remarkable human beings: driven, insightful, and, above all, compassionate. I could not have asked for better advisors.

I am equally thankful to my thesis committee – Adam Elmachtoub, Nikhil Garg, and Will Ma – for their thoughtful feedback and encouragement. Will went a step further, serving as a letter writer during my job market and offering the kind of informal mentorship that quietly shapes a career; his example remains a source of inspiration.

Beyond the committee, the Columbia has been a remarkable place to learn. I have benefited immensely from conversations with Nick Arnosti, Santiago Balseiro, Carri Chan, Jing Dong, Paul Glasserman, Vineet Goyal, Henry Lam, Hannah Li, Hong Namkoong, Hongyao Ma, Tianyi Peng, Daniel Russo, and Assaf Zeevi. Each has influenced my thinking in ways large and small. I am also thankful to Elizabeth, Andrew, Cristina, Maria, Winnie and Samantha for ensuring that I have a smooth time at the department.

My academic path began long before Columbia, and I am grateful to mentors who set me on it: Ranveer Chandra, Vijay Subramanian, and Rahul Vaze. Their belief in me during my undergraduate and master's years gave me the confidence to pursue doctoral study.

A special thanks goes to the Amazon F2P team – Andrea Qualizza, Doug Wines, Jikai Zou, and Weihong Hu – for a summer that made me more independent and taught me the invaluable lesson of breaking things and moving fast while keeping an eye on real-world impact.

Research is seldom a solitary endeavor, and my life at Columbia was enriched by countless conversations (and famously long lunches) with fellow students: Alfred, Anand, Boris, Daksh, David, Ethan, Harsh, Jerry, Omar, Prakirt, Priyank, Rachitesh, Sasank, Sastry, Shawn, Utkarsh, Wen, and Yuri. Your camaraderie made even the toughest stretches feel lighter. I was also fortunate to collaborate with Yilun Chen and Wenxin Zhang; their insight and persistence made collaboration an absolute pleasure.

My transition to the United States was eased immeasurably by my aunts (Jyanti and Jyoti mausi), uncles (Raj and Satyen mausa), and cousins (Yash, Utkarsh and Sudeeksha) here, who opened their homes and hearts to me. You were my family away from home, and your warmth turned a new country into familiar ground.

To my parents, Naresh and Jayashree: every achievement of mine is built upon your love and sacrifices. To my brother Pradyut, thank you for being a confidant and unfailing source of support.

Finally, to Ganga, my partner and my rock: your patience, steadfast encouragement, and unshakeable belief in me carried this thesis to the finish line. You celebrated the highs and steadied me through the lows; without you, none of this would have been possible.

To all of you—named and unnamed—thank you. This thesis is as much yours as it is mine.

## **Dedication**

*To my parents, my brother, my family and Ganga*

## Executive Summary

In today's interconnected world, online platforms have become pivotal in shaping economic and social landscapes. This transformation underscores the critical need to understand and optimize the operations of these platforms. A key operational task that these platforms perform is that of *matching* and this has been the central theme of this thesis. Examples range from order fulfillment on e-commerce platforms like Amazon and Walmart, connecting service providers and customers on online labor marketplaces like Handy and Upwork, to recommending content on media streaming platforms like Netflix and Spotify.

Despite the prevalence of these platforms, there remains a substantial gap between current practices and efficient operations, potentially undermining the ability of these platforms and marketplaces to deliver on their promise. This gap exists primarily for two reasons: (i) there can be a poor understanding of the drivers of (in)efficiencies in these systems, and (ii) implementing the optimal solution is often complex and practically infeasible. These factors, either individually or collectively, can contribute to less-than-ideal outcomes, such as inefficiencies in supply chain operations or suboptimal user experiences on media streaming platforms. We seek to narrow these gaps by interweaving two main pillars:

- (1) *Model and understand fundamental drivers of (in)efficiencies.* On the diagnostic end, we focus on identifying key issues in the operations of online platforms. By developing models that capture the specific facets and tensions within these contexts – such as the interplay between the distribution of request types and algorithmic performance in the context of dynamic resource allocation problems or impact of optimizing proxy metrics in the context of recommendation systems – we provide a nuanced understanding of how these systems operate. This systematic assessment helps identify the underlying drivers of performance, offering novel and valuable insights that can inform more effective design and operations of these platforms.
- (2) *Develop methods to optimize operations.* On the prescriptive end, we design algorithms and

tools to enhance the operations of these platforms. In particular, the emphasis is on developing *simple* algorithmic principles so as to provide practitioners with actionable insights.

The dissertation is organized into five self-contained but thematically linked chapters – the first three chapters are related to problems in online matching, dynamic resource allocation and multi-objective optimization in order fulfillment problems, while, the last two chapters deal with design and optimization of recommendation systems. Below, we present an outline of the thesis along with an executive summary of each chapter.

## **0.1 Online matching and Dynamic Resource Allocation: Models and Algorithms**

In the first part of the thesis, we focus on dynamic resource allocation problems such as online matching and order fulfillment.

### **0.1.1 Chapter 1: Feature-Based Dynamic Matching**

**Problem Motivation.** Matching platforms that operate in a centralized manner face a significant challenge in dynamically assigning supply units to arriving demand units. In such platforms, both customers (demand) and service providers (supply) are highly heterogeneous, and the pool of service providers is often limited. In assigning a service provider to fulfill each incoming request, the platform faces a trade-off between better serving the current customer request and preserving highly valued service providers for future customers.

**Model.** We introduce a stylized model that captures the aforementioned key factors of a centralized matching platform. We use ( $d$ -dimensional) vectors to characterize the heterogeneous demand and supply, where each dimension represents a particular feature such as demographic information, target type of services, acceptable price (for customers) and expertise, ratings, cost (for service providers). Customers characterized by i.i.d. demand weight vectors dynamically arrive on the platform and request an immediate match to a service provider, where a pool of service providers, each characterized by an i.i.d. supply feature vector, is initially available at the outset.

The platform must decide whether to assign a service provider to fulfill the demand or reject the demand immediately and irrevocably upon the arrival of a customer. Once a service provider is assigned, she leaves permanently, depleting the supply pool by one. Our model highlights the market heterogeneity by incorporating a demand weight vector and supply feature vector dependent matching utility function, thus certain customers are better served by certain service providers, depending on their particular types. The platform has a centralized objective: to maximize the expected average matching utility generated.

**Contribution.** Our main contribution is the design and the analysis of a principled algorithmic approach to the aforementioned operational challenge, dubbed **Simulate-Optimize-Assign-Repeat (SOAR)**, which combines practical implementability and a strong theoretical near-optimality guarantee. The key idea of **SOAR** is to utilize the power of simulation to facilitate efficient online decision-making. More precisely, at each time period  $t$ , we simulate a set of  $n - t$  future demand units. Together with the realized time  $t$  demand unit, they form a projected demand pool. We compute the optimal offline assignment between the projected demand pool and the remaining supply pool and match the time  $t$  demand as per its match under this assignment. We show that **SOAR** enjoys a surprisingly universal near-optimal performance guarantee across different modeling assumptions. En route to proving our guarantees, we develop a general performance analysis framework, which draws a novel connection between the performance of **SOAR** and the sequence of hindsight optima of the matching problem, which may be of wider applicability and independent interest.

**Insights.** Our results reveal a number of interesting insights. First, the "cost of matching" dominates the "cost of uncertainty about the future" in our setting. In particular, **SOAR** attains the same regret scaling as the hindsight optimal matching in all cases, indicating that knowing the future demand in advance does not substantially boost matching performance. Second, as  $d$  increases, matching becomes harder. Since  $d$  corresponds to the level of heterogeneity on both sides of the market in our model, this observation can be interpreted as matching becoming harder in a more

heterogeneous market, which is intuitive.

### 0.1.2 Chapter 2: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances

**Problem Motivation.** Online resource–allocation problems show up whenever a platform must dynamically allocate a finite set of resources to a stream of requests – shipping stock from many warehouses to thousands of zipcodes, assigning impressions to advertisers, or deciding which passengers board a flight. Existing theory has treated two idealised worlds: (i) a *small*, discrete set of request types (yielding constant–regret algorithms) and (ii) a smooth continuum with no gaps (where logarithmic regret is attainable). E-commerce fulfillment, however, falls in neither camp: demand locations number in the tens of thousands and are *clustered*, leaving large regions with no demand at all. This observation raises critical questions: *Which feature of the request–type distribution truly dictates the attainable regret? What principles lets an online policy attain that regret? Can we design a single, practical algorithm that works near–optimally whatever the distribution?*

**Contribution.** We first focus on the classical *multi-secretary problem*: a decision maker may hire up to  $B$  out of  $T$  sequentially arriving candidates whose abilities are drawn i.i.d. from a known distribution  $F \subset [0, 1]$ . Regret is measured against the clairvoyant policy that sees every candidate in advance. We show that performance is governed by a single parameter  $\beta$ , capturing how sharply probability mass accumulates next to *gaps* – intervals of zero density – in the support of  $F$ . For the broad class of  $(\beta, \varepsilon_0, \delta)$ –*clustered* distributions, any online policy suffers at least  $\Omega(T^{\frac{1}{2} - \frac{1}{2(1+\beta)}})$  regret. The spectrum interpolates between the constant regret discrete world ( $\beta = 0$ ) and the logarithmic continuum ( $\beta \rightarrow \infty$ ), showing that local sparsity, not mere discreteness, drives performance. The standard certainty equivalent (CE) policy fails in the presence of a gap. We propose *Conservativeness with respect to Gaps* (CWG): whenever the CE threshold is close to the edge of a gap, snap it to the edge itself. The resulting algorithm CWG matches the lower bound (up to polylog) for every  $\beta$ . Building on the single-sample simulate and optimize idea of

Chapter 1, we lift CWG to general resource allocation problems. The algorithm *Repeatedly Act using Multiple Simulations (RAMS)* operates as follows: at each epoch it simulates several future demand scenarios, evaluates the hindsight reward of each feasible action under those scenarios, and chooses the action with the highest average. The meta-theorem shows that RAMS inherits the best regret guarantee of any base algorithm satisfying mild conditions, making it near-optimal for multisecretary, network revenue management, and fulfillment instances *without tuning*.

### 0.1.3 Chapter 3: MOTIF: Multi-Objective Tradeoff In Fulfillment

**Problem Motivation.** We worked with the order assignment team of a large online retailer and marketplace to improve their order assignment engine (OAE). The core model of OAE is a Lexicographic Goal Programming Mixed Integer model incorporating multiple objectives, in use since 2012. When engine launched, it optimized only a few objectives and, for many years, achieved outstanding speed-and-cost performance. Over time OAE enriched the model to keep up with the growing business needs and increasing complexity, such as adding delivery synchronization for in house logistics and 3rd party companies, additional cost considerations (e.g., opportunity costs, load balancing). In 2021-2023, specifically OAE introduced Cost Relaxation (CR), a variant of goal programming, that was a key enabler of network regionalization in 2023 and the associated speed and cost-to-serve benefits it brought. However, with online retailer’s large scale, the system quickly becoming unweildy and highly inefficient – even *small* inefficiencies at the per order level balloon to *large* inefficiencies at the online retailer’s scale. This necessitated a complete overhaul of the system.

**Solution.** We developed a replacement to Cost Relaxation, dubbed MOTIF, which restored Pareto efficiency in OAE by employing a blended/weighted objective (BO) approach. MOTIF’s design is grounded in the idea that while at the micro-level (per order level), the optimization problem is nonconvex, at the macro-level (across millions of orders) the problem is approximately convex (due to Shapley-Folkman Theorem [1, 2]) – this necessitated a critical mindset shift from

a combinatorial viewpoint to a convexity viewpoint. In addition to this foundational different approach, we make additional engineering innovations to operationalize and scale MOTIF: (i) near real time generation of Pareto frontiers using the augmented  $\varepsilon$ -constraint method to enable business leaders to choose their preferred operating regimes under different network conditions and (ii) symmetry reduction techniques to reduce the number of decision variables, minimize solve time and reduce millions of MIP solves per day.

**Result & Impact.** The blended objective approach made the tradeoffs between different objectives consistent and enabled better control, allowing OAE to adapt to various network conditions. The BO solution resulted in Pareto improvements over CR. MOTIF ..

1. improved in-region assignments and consolidation (units per box) while reducing cost per unit.
2. significantly reduced computational overheads and scaled seamlessly with network growth thanks to the modular approach.

## 0.2 Design of Recommendation Systems: Models and Insights

Next, we take the perspective of a system designer and tackle platform design questions in the context of recommendation systems.

### 0.2.1 Chapter 4: On the Perils of Optimizing the Measurable

**Problem Motivation.** Recommendation systems are pivotal in shaping user experiences across digital platforms, yet their propensity to prioritize popular content over niche items raises concerns about exploration and user utility. While these systems aim to uncover hidden gems, their reliance on engagement metrics (e.g., clicks, watch time) often leads to a “popularity bias,” where widely consumed items dominate recommendations. This bias stems from a misalignment between engagement signals and true user utility, as engagement metrics fail to capture the intrinsic value users derive from content [3, 4]. Existing work highlights this tension but lacks structural insights

into balancing exploration and exploitation. Our study addresses this gap through a theoretical and numerical analysis of recommendation policies, focusing on their implications for engagement and utility.

**Model.** We model a platform recommending two item types: popular (P), with fixed utility, and niche (N), with zero-mean utility distributed heterogeneously across users. Niche items yield high utility for a small fraction of users (parameterized by  $p$ ) but low utility for most, reflecting real-world variability. The platform’s objective – maximizing engagement (short-term clicks) versus utility (long-term value) – creates a critical trade-off.

**Contribution.** Through theoretical analysis, we demonstrate that engagement-maximizing policies (e.g., recommending only popular items) suppress exploration of niche items, leading to homogeneous recommendations and suboptimal utility. In contrast, our utility-aware heuristic PEAR, which balances popular and niche recommendations, achieves significantly high utility with only minimal engagement loss for forward-looking platforms. This asymmetry underscores the potential to enhance user value without sacrificing engagement, challenging the status quo of popularity-driven algorithms. To generalize these insights, we relax prior utility knowledge assumptions in numerical experiments, modeling niche utility via Pareto distributions. We compare APP (engagement-optimal, recommends only popular items) with DICE (exploration-driven, mixes popular and niche initially). Results show that while APP marginally outperforms DICE in engagement (especially for lighter-tailed distributions), DICE significantly improves utility by up to 50%—highlighting the viability of exploratory policies even without explicit utility measurement. These findings persist across distributional assumptions, reinforcing the robustness of exploration for utility maximization. Our contributions are threefold: (i) *theoretical evidence* of structural misalignment between engagement and utility optimization, (ii) *quantification* of asymmetric trade-offs favoring utility-aware policies, and (iii) *numerical validation* of heuristic strategies under general settings.

**Insights.** This work advocates for rethinking recommendation design, emphasizing that modest exploration can reconcile engagement goals with substantial utility gains, ultimately enriching user experiences.

## 0.2.2 Chapter 5: Impact of Rankings and Personalized Recommendations in Marketplaces

**Problem Motivation.** Individuals make decisions—from routine choices like picking a movie to pivotal ones like selecting a college—often without full information. Many later express regret, as evidenced by surveys and personal experience. To mitigate such uncertainty, public rankings (e.g., U.S. News & World Report, IMDb) aggregate a notion of “average quality.” While these rankings guide some users effectively, they often fail to account for individual preferences. By contrast, personalized recommendation systems (e.g., Netflix) tailor options to each user’s tastes, yet remain underutilized in capacity-constrained settings like college admissions or online marketplaces. This raises our research question: *What are the implications of different information provisioning tools, such as public rankings and personalized recommendations, in environments with and without supply-side constraints?*

**Model.** Let  $\mathcal{X}$  be the set of agents and  $\mathcal{Y}$  be the set of items, with  $|\mathcal{Y}| = n$ . Each agent chooses *one* item. In uncapacitated settings, multiple agents can pick the same item. In capacitated settings, each item has unit capacity, so exactly  $n$  agents are matched to  $n$  items. Each agent  $x$  derives utility  $u_{xy} = (1 - \rho) q_y + \rho \varphi_{xy}$  from item  $y$ . The term  $q_y$  (“common quality”) is item-specific and captures an overall measure of quality. The term  $\varphi_{xy}$  (“idiosyncratic preference”) captures personalized fit. The parameter  $0 \leq \rho \leq 1$  determines how heterogeneous preferences are: higher  $\rho$  means  $\varphi_{xy}$  dominates decisions. We assume  $(q_y)$  and  $(\varphi_{xy})$  are drawn from distributions with Pareto-like or exponential tails, motivated by empirical evidence in creative and academic domains. We study the following information regimes: (i) No Information ( $\emptyset$ ): agents pick randomly, (ii) Only Quality Information ( $q$ ): agents only know  $q_y$  (e.g., a ranking) and (iii) Full Information ( $u$ ): agents know both  $q_y$  and  $\varphi_{xy}$  (e.g., personalized recommendations).

**Insights.** Our main findings in the uncapacitated and capacitated settings are as follows:

*Uncapacitated Environments.* Public rankings substantially increase welfare when  $\rho$  is small, as most agents benefit from knowing the shared quality  $q_y$ . Personalized recommendations yield *additional* improvements by incorporating individual preferences  $\varphi_{xy}$ , which is especially valuable when  $\rho$  is large (highly heterogeneous tastes).

*Capacitated Environments.* Under capacity constraints, public rankings alone do not improve aggregate welfare. High-ranked items become overly sought after, creating congestion and leaving some agents worse off. However, personalized recommendations enable a more efficient allocation of items by matching agents to the options they value most. This avoids under-matching, prevents excessive competition for a single top choice, and can significantly increase overall utility.

# Chapter 1: Feature-Based Dynamic Matching

*Based on the paper [5] co-authored with Yilun Chen, Yash Kanoria and Wenxin Zhang.*

## 1.1 Introduction

Centralized matching platforms have transformed the way customers access services in sectors such as hospitality, transportation, and finding a job. These platforms facilitate on-demand connections between customers and service providers, based on customer needs and service provider attributes and availability. A common feature of many of these platforms is that the demand and supply sides are often *both* highly differentiated. For instance, on online home services platforms such as Handy.com, customers arrive and specify various service requests (e.g., plumbing, room cleaning, furniture repair) and personalized service preferences (location, time, price). Simultaneously, service providers on these platforms are differentiated by their diverse skill sets, locations, and availability schedules. The value of a match between a given customer and a service provider depends on both the specific preferences of the customer and the attributes of the service provider. Platforms are tasked with maximizing the overall matching value of customers, a composite measure that incorporates customer satisfaction, service provider value, and the platform's value, for those arriving over a finite horizon.

The varied nature of both demand and supply poses challenges to the platform's decision-making, especially when supply is constrained. With a limited pool of service providers and the arrival of customers over time, the platform faces a trade-off between (a) optimizing the value of the current match and (b) maintaining a diverse pool of service providers for future customers. While prioritizing (a) offers maximum immediate value, implementing (b) may bring the benefit of facilitating higher valued matches for future customers. Balancing these two objectives to ensure

optimal outcomes is, in general, not straightforward. This motivates our central research question:

*In a dynamic two-sided matching setting with heterogeneous customer and service providers, how should a centralized platform match customers arriving over time to maximize the overall value generated?*

We capture the platform’s problem through a *feature-based* dynamic matching model. In our base model, a pool of  $n$  service providers is initially available and  $n$  customer requests arrive sequentially. Upon the arrival of each customer request, the platform immediately and irrevocably assigns to it an available service provider, who then leaves the pool permanently. In our model, customers and service providers are characterized by multidimensional vectors, where each dimension numerically encodes an attribute observable to the platform, for instance, job location, listed prices, ratings, expertise (for service providers), and demographics in general. We capture market uncertainty and heterogeneity by assuming the demand weight vectors (that represent customers) and supply feature vectors (that represent service providers) are drawn *i.i.d.* from some demand distribution  $P$  and supply distribution  $Q$ , respectively. The match quality (or match value) between a customer and a service provider is modelled by a quality function  $\varphi(X, Y)$  that depends on both of their weight and feature vectors  $X$  and  $Y$  respectively, one typical example being  $\varphi(X, Y) = \langle X, Y \rangle$ . The platform’s goal is to repeatedly make matching decisions to maximize the expected average matching quality. In this paper, we also consider the setting of scarce supply where the number of customers is more than the number of service providers. We show that our modelling framework is flexible enough to incorporate such a setting (refer to Section 1.2.1).

**Modelling innovation.** Much of the prior research on stochastic dynamic two-sided matching and resource allocation assumes a small number of demand and supply (or resource) types ([6, 7, 8, 9]). Such an assumption is critical in ensuring theoretical guarantees but leads to a limitation in modeling capability. In contrast, our feature-based modeling framework allows for many or even an infinite number of demand and supply types while remaining analytically tractable. This is achieved by exploiting proper continuity conditions such as a spatial structure on the feature

vector spaces and the matching quality functions. In particular, we obtain algorithms that provide provably asymptotically optimal performance guarantees, i.e., our performance loss relative to the limiting hindsight optimal solution vanishes as  $n$ , the number of supply (demand) units, increases.

Within the modeling framework, we study and quantitatively compare different matching policies. While it may be tempting for practitioners to greedily assign service providers that maximize immediate matching quality, we observe that such myopic policies may end up incurring significant quality loss, and are thus highly sub-optimal (see Proposition 1). The sub-optimality stems from overlooking future customers and exhausting a particular group of similar service providers too early. This observation necessitates the design of forward-looking algorithms.

### 1.1.1 Main Contribution

In this work, we develop a simple forward-looking algorithm dubbed **SOAR**, which automatically preserves supply diversity and achieves near-optimal performance within our model across a variety of settings. We also develop an analytical framework for our algorithm **SOAR** which in turn enables us to prove near-optimal performance guarantees and may be of broader interest. As a corollary of our techniques, we also resolve some open problems posed in [10]. We now elaborate on our contributions.

- (i) *A simple forward-looking algorithm.* We propose a principled approach dubbed **SOAR** (short for **Simulate-Optimize-Assign-Repeat**), that combines real-world applicability and a theoretical performance guarantee. **SOAR** is inspired by model predictive control (MPC), a well-known heuristic in dynamic control theory (see e.g. [11, 12] for detailed discussion). The key idea is to utilize a simulated scenario of future demands to facilitate efficient online decision-making. More precisely, at each decision epoch (namely, when a new customer arrives), the algorithm forms a projected demand pool, which includes both the currently arriving customer and a simulated stream of future customers. It then matches the arriving customer as per their assignment under the optimal offline matching between the projected demand pool and the remaining supply pool. The two main steps of **SOAR** at each deci-

sion epoch are (i) simulating a future demand stream and (ii) solving an offline assignment problem, both easy to implement. Regarding (i), we note that many companies collect and store large amounts of demand data. This rich data can be leveraged to build high-fidelity simulators for SOAR. Furthermore, our algorithm only requires simulating one sample path of future customers in each decision epoch, setting it apart from most simulation-based algorithms which fundamentally rely on sample average approximation (SAA) and require simulating a large number of independent streams of future customers. Regarding (ii), the bipartite assignment problem is a classical computational problem. Theoretically, a worst-case guarantee of  $O(n^{2+o(1)})$  has been shown for this problem [13], ensuring its tractability. In practice, several fast and highly scalable solvers have been developed for this problem that can be directly plugged into SOAR.

(ii) Novel analytical framework for SOAR. We note that MPC is generally considered to be a sub-optimal heuristic (see Section 6.5 of [11]). Our main technical contribution lies in developing a novel framework that allows us to systematically analyze and establish the near-optimal performance of SOAR for stochastic feature-based dynamic matching. Serving as the technical backbone of this work, our framework relies on a key observation (Theorem 1), which expresses the performance of SOAR as the average of a sequence of hindsight optimum values. In particular, when there are  $n$  supply providers and customers, let  $U_n(\text{SOAR})$  denote the expected average matching quality under SOAR and  $U_n^{\text{H}}$  denote the expected average matching quality achievable in hindsight. Then Theorem 1 asserts that  $U_n(\text{SOAR}) = \frac{1}{n} \sum_{k=1}^n U_k^{\text{H}}$ . Equivalently in terms of regret, we have  $\text{REG}_n(\text{SOAR}) = U_\infty - U_n(\text{SOAR}) = \frac{1}{n} \sum_{k=1}^n (U_\infty - U_k^{\text{H}})$ , where  $U_\infty$  denotes the thick market limit (refer to Section 5.2). This result leverages the symmetry induced by *i.i.d.* random variables, and is otherwise completely general; it holds for arbitrary demand and supply distributions  $P, Q$ , and for arbitrary matching quality function  $\varphi(X, Y)$  that is bounded over the support of  $X$  and  $Y$  (boundedness is required only to ensure these performance metrics are well-defined). It reduces bounding the regret of the *online* algorithm SOAR to character-

izing a sequence of bounds on the regret under the hindsight relaxation for  $k = 1, 2, \dots, n$ , which are purely *offline*. We leverage this framework to prove near-optimal regret guarantees for different quality functions and under general structural assumptions on the demand and supply distributions.

- (iii) Near-optimal performance of SOAR. SOAR enjoys a guarantee of near-optimal performance across broad classes of demand and supply distributions and quality functions. Our general performance guarantee, Corollary 2, states that SOAR is near optimal for any matching problem satisfying a “regular scaling” property (Definition 1), which we believe essentially incorporates all non-pathological matching instances. In particular, we demonstrate that under a specific performance measure, namely, the average matching regret relative to the hindsight limit matching value, the performance of SOAR is within a  $\log n$  factor of that of the hindsight optimum for any matching instance which scales regularly. We then explicitly characterize the scaling of average matching regret incurred by SOAR in various interesting classes of matching problems. In Section 1.4, we investigate matching for a general class of quality functions  $\varphi_p(X, Y) = -\|X - Y\|^p$  ( $\|\cdot\|$  denotes Euclidean norm) for  $p \geq 1$  under two sets of assumptions on the demand and supply distributions, one restricting  $P$  and  $Q$  to be the uniform distribution over  $[0, 1]^d$ , following the literature on dynamic spatial matching ([14, 10, 15]), and the other allowing for arbitrary distributions. We characterize the regret scaling attained by SOAR under the two sets of assumptions respectively, which are both optimal (up to a factor of at most  $\log n$ ) for each possible dimension  $d$  and exponent  $p$  (see Theorem 2). In particular, sharper regret scaling is achievable under the more restrictive assumption that both distributions  $P$  and  $Q$  are uniform over  $[0, 1]^d$ . In the same section, we also study the practically relevant dot-product matching quality function  $\langle X, Y \rangle$ . In particular, we establish an equivalence between the dot-product quality function and  $\varphi_p$  with  $p = 2$ , and prove in this special case that SOAR attains the sharper regret scaling for a general class of smooth, regular, “uniform-like” distributions  $P$  and  $Q$  that can be non-uniform and unequal (see Corollary 3, Theorem 3, and Table 1.1). The matching upper and lower bounds in

Table 1.1 together characterize the complete landscape of the regret scaling with dot product quality. As a corollary of our analysis, we also solve an open problem in [10] and generalize some of the results in [10] to a setting with quality function  $\varphi(X, Y) = -\|X - Y\|^p$  for  $p > 1$ . Our results significantly advance the previous understanding of dynamic spatial matching, which is mostly restricted to the special case with  $P = Q = \text{Uniform}([0, 1]^d)$  and  $p = 1$ , [10, 15, 14].

Our results reveal several interesting insights as we summarize below. First, the “cost of matching” dominates the “cost of uncertainty about the future” in our setting. In particular, **SOAR** attains the same regret scaling (up to logarithmic factors) as the hindsight optimal matching in all cases, indicating that knowing the future demand in advance does not allow us to substantially reduce regret. Second, the regret of **SOAR** increases as the dimension of vectors,  $d$ , increases. Since  $d$  corresponds to the level of heterogeneity on both sides of the market in our model, this observation can be interpreted as matching is harder in a market with more dimensions of heterogeneity. Finally, smoothness helps reduce regret, but only in low dimensions. For  $d \geq 4$  the regret scaling achievable is the same under smooth versus arbitrary distributions. Section 1.5 describes a range of numerical studies, which provide evidence that **SOAR** is near optimal and exhibits regret which vanishes rapidly with  $n$  in various settings.

Table 1.1: The average regret scaling for  $\varphi(X, Y) = \langle X, Y \rangle$ . Here  $a \wedge b := \min\{a, b\}$ .

	$P = Q = \text{Uniform}([0, 1]^d)$	$P, Q \text{ Smooth}$	$P, Q \text{ Arbitrary}$
Lower Bound	$\tilde{\Omega}(n^{-(\frac{2}{d} \wedge 1)})$	$\tilde{\Omega}(n^{-(\frac{2}{d} \wedge 1)})$	$\tilde{\Omega}(n^{-(\frac{2}{d} \wedge \frac{1}{2})})$
Algorithm of [10]	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d} \wedge \frac{1}{2})})$	-	-
<b>SOAR (this work)</b>	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d} \wedge 1)})$	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d} \wedge 1)})$	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d} \wedge \frac{1}{2})})$

### 1.1.2 Related Literature

*Matching Markets.* Devising algorithms and guarantees for online matching is a central topic

for researchers at the intersection of CS, Economics, and OR/OM; see, e.g., [16, 17, 8, 18, 19, 20, 21, 22] among many others. In recent years amid the rise of the platform economy, matching markets, with a rich set of newly emerging economical/operational/computational challenges, have attracted significant attention in the academic literature [23, 24, 25, 26, 27, 28, 29, 30]. In this paper, we study a centralized online platform that aims at maximizing social welfare (expected total matching quality). Like us, several works consider a reward-maximizing platform, where the reward can be total revenue, total match number, etc. [30] studies a centralized matching problem in a stochastic environment with departures and formulate it as an MDP. [26] and [29] investigate a decentralized platform, and how to choose a good market equilibrium through appropriately designed match recommendations. [25] conducts an insightful study revealing *why* platforms in the home services industries tend to implement centralized matching policies, complementary to our work.

*Dynamic Resource Allocation.* We formulate and solve an online matching problem with *i.i.d.* demand and supply, which is categorized as a dynamic resource allocation problem in the OR/OM literature. The  $d = 1$  special case of the model has appeared in earlier works in the OR/OM community under the name of (stochastic) sequential assignment problem ([31, 32], with applications in the kidney exchange markets [33]). The workhorse policy commonly used in several classical dynamic resource allocation instances is the **Certainty Equivalent (CE)** policy ([34]). It is worth pointing out that directly applying CE in our setting requires solving a fluid problem equivalent to an optimal transport (OT) problem, which can be challenging when  $P$  and  $Q$  are supported on infinite or even uncountable sets ([35]). Indeed, the finite-ness of the support of demand/supply distributions is a key requirement not only to the design of CE but also to the analysis of several new algorithms recently proposed in the field ([36, 37]). By comparison, **SOAR** is a simulation-based computationally efficient proxy for CE with performance guarantees, that naturally copes with infinite types of demand and supply in the matching context. We note here that simulation-based policies have previously been proposed and studied in the context of dynamic resource allocation problems [38, 39, 40, 41, 42, 43]. The most relevant to our work is [43]. Subsequent to our work,

in a recently revised version of [43], the authors study a multi-simulation variant of our SOAR-like policy. They use the compensated coupling framework of [36] to analyze hindsight-based regret, and their theoretical guarantees on the dynamic matching problem are only applicable under the setting with a few demand and supply types. Different from previous work, our performance guarantee does not rely on sample-average-approximation (SAA) type analysis, which depends on the number of simulated sample paths used. Instead, the SOAR algorithm developed in this paper mimics MPC and only uses one simulated sample path in each decision epoch. We develop a formula that characterizes the expected performance of SOAR (cf. Theorem 1), enabling us to prove optimal regret scalings in various settings for very general demand and supply type distributions.

*Stochastic Online Matching.* The problem studied in this paper is relevant to the literature on stochastic online matching. This literature examines the stochastic online bipartite matching problem in the known *i.i.d.* input model, initiated by [44]. Many subsequent papers have generalized the base model of [44], which assumed an unweighted underlying graph and integral arrival rates, and proposed new algorithms. In particular, [45], [46], and [47] derived algorithms with improved competitive ratio guarantees and studied models that relaxed the integrality restriction on the arrival rates, allowing the arrival rates to be arbitrarily close to zero. Additionally, [48] and [47] extended the base model to the edge-weighted setting. A key difference between our setting and the existing literature is the power of the adversary. In the stochastic online matching literature, such as [44] and its successors, the adversary can choose the supply units arbitrarily, leading to hard instances with non-vanishing regret. In our setting, the supply units are *i.i.d.* draws from a distribution, limiting the adversary’s power as  $n$  grows (to “selecting” the “worst-case” distributions and the matching quality function, in particular). Consequently, our setting inherently enjoys vanishing regret, regardless of regularity assumptions like the continuity or smoothness of  $P, Q, \varphi$  (see Corollary 1). Subsequent to our work, [49] studies a stochastic online metric matching problem where the adversary chooses supply unit locations in a metric space. They leverage the algorithmic analysis framework developed in this paper and achieve improved competitive ratio and provably near optimal regret.

*Dynamic Spatial Matching.* An intimately related literature to this work is dynamic spatial matching. There the matching cost is typically chosen as the distance between two units. The stochastic version of dynamic spatial matching is a specific instance in our model, with  $P = Q = \text{Uniform}([0, 1]^d)$  and  $\varphi(x, y) = -\|x - y\|$ . [10] studies this problem (under the so-called “semi-dynamic setting”) and gives a complete characterization of the regret scaling. Their Hierarchical Greedy algorithm achieves optimal scaling for all  $d \geq 1$ . However, the results and techniques do not extend to general quality  $-\|x - y\|^p$  and unequal demand and supply distributions as we previously alluded to. SOAR resolves an open problem posed in [10] by achieving a provably tight regret scaling under  $-\|x - y\|^p$  and  $P = Q$ ; see Proposition 2. Other related works include [50] which proposes a gravitational allocation method for spatial matching in both dynamic and offline settings between two sets of uniform points on a  $3D$  sphere that achieves tight guarantee. [14], [15], and [10] all study the case of excess supply, whereas we focus on the case of scarce supply in this work. In general, randomized algorithms that are in spirit similar to MPC have been proposed to solve online matching problems (e.g., [51]), that come with different performance guarantees (e.g. competitive ratio) under various modeling assumptions. We believe that an attractive feature of our work relative to prior work is the broad generality of the conditions under which we establish near optimality of SOAR.

### 1.1.3 Notation

We denote  $f(n) = \Theta(g(n))$  if there exists constant  $C > 0$  independent of  $n$  such that  $C^{-1}g(n) \leq f(n) \leq Cg(n)$ . Similarly, we say  $f(n) = \mathcal{O}(g(n))$  if there exists constant  $C < \infty$  independent of  $n$  such that  $f(n) \leq Cg(n)$  and we say that  $f(n) = \Omega(g(n))$  if exists a constant  $C > 0$  independent of  $n$  such that  $f(n) \geq Cg(n)$ . The notation  $\tilde{\mathcal{O}}, \tilde{\Theta}, \tilde{\Omega}$  will ignore the polylogarithmic factors in  $n$ . For example  $\mathcal{O}(n \log n) = \tilde{\mathcal{O}}(n)$ . We denote  $a \wedge b := \min\{a, b\}$ ,  $a \vee b := \max\{a, b\}$ .

## 1.2 Model

Let  $Y_1, Y_2, \dots, Y_n$  denote an initial endowment of  $n$  supply units, where each  $Y_k \in \mathcal{Y} \subseteq \mathbb{R}^d$  is sampled *i.i.d.* from a supply feature distribution  $Q$ . A sequence of  $n$  demand units arrive sequentially over time, each seeking one supply unit. The  $t^{\text{th}}$  demand unit for  $t = 1, 2, \dots, n$  is a vector  $X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ , which is *i.i.d.* drawn from a demand weight distribution  $P$ . Upon the arrival of the demand unit  $X_t$ , the platform must immediately and irrevocably assign an available supply unit  $Y$  to the current demand unit  $X_t$ . Such an assignment decision generates a feature and weight-dependent match value of  $\varphi(X_t, Y)$ , after which the matched pair leaves the system, depleting the available supply units by one. We refer to the function  $\varphi : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  as the (match) *quality* function. The above decision-making process is repeated  $n$  times, until there is no supply unit left.

The platform seeks a dynamic matching policy that maximizes the average expected match value. Let  $\mathcal{H}_t$  denote the history of the system up to decision epoch  $t$  (namely, when the  $t^{\text{th}}$  customer arrives) that incorporates the first  $t - 1$  demand units and their corresponding matched supply units. Let  $\mathcal{S}_t$  denote the set of indices for the remaining supply units when the  $t^{\text{th}}$  demand unit arrives, and  $\Delta(\mathcal{S}_t)$  be the set of probability distributions over  $\mathcal{S}_t$ . We formally define a dynamic matching policy as a collection of mappings  $\pi := (\pi_t)_{1 \leq t \leq n}$ , where  $\pi_t(X_t, \mathcal{H}_t) \in \Delta(\mathcal{S}_t)$  is the possibly randomized assignment of supply unit given the  $t^{\text{th}}$  demand unit  $X_t$  and the current history  $\mathcal{H}_t$ . We denote the average expected match value under  $\pi$  by

$$U_n(\pi; P, Q, \varphi) := \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \varphi(X_t, Y_{\pi_t}) \right], \quad (1.1)$$

where the expectation is taken over  $\pi, P, Q$ , and we slightly abuse notation  $\pi_t$  to represent both the random variable and its associated distribution.

We do not assume the platform knows  $P$  and  $Q$ . Instead, we assume that the platform has access to  $m$  *i.i.d.* samples from  $P$ , collected from its historical matching activities. Our algorithm and the corresponding theoretical results only require a moderate  $m$  relative to the scale of the problem. In particular,  $m = \Omega(n^2)$ .

**Benchmark.** We use the limiting hindsight optimum of the problem as a benchmark to measure and compare the performance of matching policies. To define this benchmark, we first introduce the following hindsight relaxation

$$U_n^H(P, Q, \varphi) := \frac{1}{n} \mathbb{E} \left[ \sup_{\tau \in S_n} \sum_{t=1}^n \varphi(X_t, Y_{\tau_t}) \right], \quad (1.2)$$

where  $S_n$  is the symmetric group on  $\{1, \dots, n\}$ . We refer to  $U_n^H(P, Q, \varphi)$  as the *hindsight optimum value* of the problem. In words, the hindsight optimum is achieved by relaxing the non-anticipative constraint on  $\pi$  and allowing access to the actual values of all arriving demand units  $X_1, \dots, X_n$  at time 0. The *limiting hindsight optimum value* of the problem, denoted as  $U_\infty(P, Q, \varphi)$ , represents the thick market limit of the hindsight optimum:  $U_\infty(P, Q, \varphi) := \lim_{n \rightarrow \infty} U_n^H(P, Q, \varphi)$ , whose existence is guaranteed by the boundedness and monotonicity of  $U_n^H$  (see Appendix A.1 for details.) For any dynamic matching policy  $\pi$  and  $n \geq 1$ , we have  $U_\infty(P, Q, \varphi) \geq U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$  (due to the monotonicity of  $U_n^H$ ; as formalized in Lemma 5 in Appendix A.1).

We take  $U_\infty(P, Q, \varphi)$  to be our performance benchmark, and define *regret* (of a policy  $\pi$ ) as

$$\text{REG}_n(\pi; P, Q, \varphi) := U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi). \quad (1.3)$$

For  $\text{REG}_n(\pi; P, Q, \varphi)$  to be well-defined, we assume the quality function  $\varphi$  is bounded. Note that  $U_\infty(P, Q, \varphi)$ ,  $U_n^H(P, Q, \varphi)$  and  $U_n(\pi; P, Q, \varphi)$  are all average (per match) quantities. We further note that under proper regularity conditions,  $U_\infty$  coincides with the optimal transport value between the distributions  $P$  and  $Q$  with respect to the function  $\varphi$  (For a counterexample, see Remark 8). We shall hereafter drop the notation for dependence on distributions  $P$  and  $Q$  and the quality function  $\varphi$ , denoting by  $U_\infty$ ,  $U_n^H$ ,  $U_n(\pi)$  and  $\text{REG}_n(\pi)$  the corresponding objectives, since the distributions are always clear from the context.

### 1.2.1 Modelling scarce supply and rejection cost

So far in our discussion, we have focused on a balanced market setting with an equal number of supply and demand units. However, our model is flexible enough to incorporate the setting with scarce supply and the corresponding rejection cost due to not being able to serve all demand units. To this end, we introduce the notion of a *virtual* or *dummy* supply unit denoted as **dum**. Matching a demand unit to a virtual supply unit **dum** is akin to rejecting that particular demand unit. To capture scarce supply within our model, we assume each supply unit is drawn *i.i.d.* from a specific distribution  $Q'$ , which is a mixture of  $Q$ , the distribution of *real* supply units, and  $Q^{\mathbf{dum}}$ , a Dirac measure on the atom **dum**. Such a distribution ensures each supply unit is virtual with a certain probability  $p$ . Consequently, the number of  $Q$ -distributed *real* supply units corresponding to  $n$  demand units follows a  $\text{Bin}(n, 1 - p)$  distribution and is always (weakly) less than  $n$ , thus capturing scarce supply. Furthermore, our model can incorporate rejection cost. Let  $c(x)$  denote the cost of rejecting a particular demand unit  $x \in \mathcal{X}$ . Then the matching quality function  $\varphi'(x, y) := \varphi(x, y)\mathbb{1}\{y \neq \mathbf{dum}\} - c(x)\mathbb{1}\{y = \mathbf{dum}\}$  effectively captures the value generated from matching the demand unit to a real supply unit and the cost incurred from rejecting the demand unit.

## 1.3 SOAR: Algorithmic Principle and Performance Analysis

The two primary properties we seek from our matching technology are: (i) computational efficiency, and (ii) provably near optimal performance. In the following discussion, we will first assess the practically popular myopic algorithm against our desiderata and argue that it can lead to highly sub-optimal matching outcomes. We then propose our algorithmic approach, SOAR, that possesses both properties, and present a meta guarantee on the performance of SOAR, which will drive the guarantees on regret scaling presented in the next section.

### 1.3.1 Insufficiency of Myopic Policies

We first consider the **Greedy** policy where each arriving demand unit is matched to a myopically optimal supply unit. In terms of computational efficiency, each matching decision can be computed in linear time. Motivated in part by ride hailing platforms, [10] and [14] show that **Greedy** has near optimal performance if the demand and supply distributions are identical (and uniform). However, this (near) optimality of **Greedy** is quite fragile. Beyond the very special cases studied in [14, 10], **Greedy** can suffer from significant performance degradation, resulting in non-vanishing regret. We formalize this using the following instance.

**Proposition 1 (Failure of Greedy)** *Suppose the supply distribution  $Q$  is supported on the atoms  $\{0, 1\}$ , i.e.,  $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = \rho$  and the demand distribution  $P$  has a continuous distribution over the interval  $[0, 1]$  with density bounded below and above, i.e., there exists  $\gamma > 0$  such that  $\gamma^{-1} \leq f_P \leq \gamma$  and has a CDF  $F_P$ . Fix  $\rho \in (0, 1)$  and assume that  $F_P(1/2) \neq 1 - \rho$ . Fix  $p \geq 1$  and consider the quality function  $\varphi(X, Y) = -|X - Y|^p$ . Then there exists a universal constant  $c = c(\rho, F_P, p) > 0$  and  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ , we have that  $\text{REG}_n(\text{Greedy}) \geq c$ .*

Note that Proposition 1 holds for a class of quality functions including the standard euclidean distance with  $p = 1$ . To obtain some intuition for this result, consider the case of  $P = \text{Uniform}([0, 1])$  and let  $\rho = 1/4$ . Consider the fluid limit of the problem. Whenever there are supply units available, the **Greedy** algorithm matches demand unit arrivals located in  $[1/2, 1]$  to the supply units at location 1 and demand unit arrivals in  $[0, 1/2]$  to supply units at location 0. However, as  $\mathbb{P}(Y = 1) = 1/4$ , the **Greedy** allocation would prematurely exhaust all the supply units at 1 by the end of the first half of the time horizon, and then for the remaining half of the time horizon, all the demand units will be matched to the supply units at 0. In contrast, the optimal fluid policy would match all the demand units located in  $[3/4, 1]$  to the supply units located at 1 and the demand units located in  $[0, 3/4]$  to the supply units located at 0. Due to the myopic nature of **Greedy**, the demand units with location in  $[3/4, 1]$  that arrive in the second half are forced to be matched to supply units at 0 resulting in large matching costs and hence the non-vanishing regret.

This intuition is formalized in our proof deferred to Appendix A.2.

### 1.3.2 Simulate, Optimize, Assign, Repeat (SOAR) Principle

We now present SOAR, a principled simulation-based approach to general stochastic dynamic two-sided matching, and establish a meta-performance analysis framework. Our key result is a weighted sum representation of the expected average matching quality of SOAR, where the terms in the summation are precisely the hindsight optimum values. Such a direct connection between the policy performance and the hindsight optimum leads to several interesting structural results, as well as concrete regret analysis in specific settings. The latter is summarized in the next section.

#### **Policy Description**

Our policy does not require the precise distributional knowledge of  $P$  and  $Q$ . Instead, we utilize independent demand samples of  $P$  to facilitate matching. For ease of mathematical exposition, we conceptualize this sampling access requirement by assuming access to a demand unit simulator  $SIM$ . More precisely, upon calling  $SIM$ , we get an independent demand sample drawn from  $P$ . A pool of  $m$  historical *i.i.d.* demand units can thus be viewed as being generated from  $m$  repeated calls to  $SIM$ .

We formally state SOAR in Algorithm 1. The algorithmic principle is simple and can be summarized as follows. Upon each arrival of a new demand, the algorithm simulates a future demand scenario and solves a hindsight assignment problem, based on which the new demand is assigned to a supply unit. It is worth mentioning that to operationalize SOAR, we only need  $\Theta(n^2)$  independent samples of demand, since only one simulated sample path and no more than  $n$  samples are needed for each demand arrival. This sets SOAR apart from many other simulation-based algorithms employing sample average approximation (SAA), which significantly relies on the volume of independently simulated samples to enhance performance. Also, there are standard algorithms, e.g., the Hungarian method and its variants, for solving the perfect assignment problem (1.4) and the state-of-the-art runtime has been reduced to almost linear time in the number of edges [13].



**Theorem 1 (Meta Performance of SOAR)** *Assuming that  $n$  supply units are drawn i.i.d. from a distribution  $Q$  and the demand units are drawn i.i.d. from a distribution  $P$ . For any quality function  $\varphi$  that is bounded on the support of  $P$  and  $Q$ , we have that the expected average match value of SOAR is given as*

$$U_n(\text{SOAR}) = \frac{1}{n} \sum_{k=1}^n U_k^H. \quad (1.5)$$

Theorem 1 holds regardless of the choice of  $P, Q$  and the quality function  $\varphi$ . The boundedness of  $\varphi$  is imposed only to ensure  $U_k^H$  is well defined. The key observation is that in each decision period, all remaining supply units, irrespective of their feature vectors, are equally likely to be assigned by SOAR to the current arriving demand unit. Such a property induces strong symmetry, and in particular, the remaining supply units continue to be distributed *i.i.d.* as per  $Q$  throughout the decision-making process, from which the theorem follows immediately.

*Proof.* The proof relies on the following key observations:

(i) *Remaining supply units are i.i.d.:* The remaining supply units at each decision epoch  $t \in \{1, 2, \dots, n\}$  are located i.i.d. according to  $Q$ .

(ii) *Expected quality is the average hindsight optimum value.* In particular,

$$\mathbb{E} \left[ \varphi(X_t, Y_{\pi_t^{\text{SOAR}}}) \right] = \frac{1}{n-t+1} \mathbb{E}_{\hat{X} \sim P, \hat{Y} \sim Q} \left[ \max_{\sigma} \sum_{k=1}^{n-t+1} \varphi(\hat{X}_k, \hat{Y}_{\sigma(k)}) \right] = U_{n-t+1}^H.$$

Note that (ii) is a corollary of (i). Indeed, upon the arrival of the  $t$ -th demand, SOAR calls an offline matching solver while outputs a perfect assignment with respect to quality function  $\varphi$  between  $\{\hat{X}_0, \dots, \hat{X}_{n-t}\}$  and  $\{\tilde{Y}_0, \dots, \tilde{Y}_{n-t}\}$ , where  $\hat{X}_0, \dots, \hat{X}_{n-t}$  is an *i.i.d.*  $P$  sequence by algorithm design, and  $\tilde{Y}_0, \dots, \tilde{Y}_{n-t}$  is an *i.i.d.*  $Q$  sequence by property (i). Hence, the expected cumulative matching quality achieved by the offline optimizer equals the hindsight optimum (cumulative) quality,  $(n-t+1)U_{n-t+1}^H$ . Due to symmetry, the expected matching quality of the matched pair  $(\hat{X}_0, \tilde{Y}_{\eta^*(0)})$  (or equivalently  $(X_t, Y_{\pi_t^{\text{SOAR}}})$ ) equals that of any pair  $(\hat{X}_i, \tilde{Y}_{\eta^*(i)})$ , from which (ii) follows.

We prove (i) inductively in  $t$ . Suppose (i) holds up to  $t - 1$ . We show that upon the arrival of the  $t^{\text{th}}$  demand, each remaining supply unit is equally likely to be matched and leave. Indeed, let  $A_i$  denote the event that  $\pi_i^{\text{SOAR}} = i$ , namely  $\tilde{Y}_i$  is matched to  $X_t$ . Let  $B(x_0, \dots, x_{n-t}, \eta)$  denote the event that  $\{x_0, \dots, x_{n-t}\}$  is the demand pool and  $\eta$  is the permutation the offline matching output by the optimizer, i.e.,  $x_i$  is matched to  $\tilde{Y}_{\eta(i)}$ . By design, one of  $x_0, \dots, x_{n-t}$  is the true realized demand of  $t$ , and the rest are the simulated units (note the random permutation of indices of the demand units in line 5 of Algorithm 1). Crucially, event  $B$  does not specify which unit is the true realized demand, and  $\eta$  is independent of which unit is the true realized demand. We thus have

$$\begin{aligned} \mathbb{P}\left(A_i | B(x_0, \dots, x_{n-t}, \eta)\right) &= \mathbb{P}\left(x_{\eta^{-1}(i)} \text{ is the true demand} \mid B(x_0, \dots, x_{n-t}, \eta)\right) \\ &= \mathbb{P}\left(x_{\eta^{-1}(i)} \text{ is the true demand} \mid \{x_0, \dots, x_{n-t}\} \text{ is the demand pool}\right) \\ &= \frac{1}{n-t+1}, \end{aligned}$$

where in the second equality we use the independence of  $\eta$  with  $\{x_{\eta^{-1}(i)} \text{ is the true demand}\}$ , and the third equality follows from the fact that the  $n - t + 1$  units in the demand pool are *i.i.d.* The above implies  $\mathbb{P}(A_i) = \frac{1}{n-t+1}$ , which then further implies property (i). Finally, Theorem 1 is an immediate corollary of (ii).  $\blacksquare$

Note that the proof of Property (i) does not require the offline matching solver to follow a particular tie-breaking rule (such as uniform-at-random tie-breaking) in facing multiple optimal solutions. Thus, Theorem 1 holds true for SOAR when implementing *any* tie-breaking rule under multiple optimal solutions to the matching problem in (1.4).

**Remark 1 (Regret Decomposition)** *Observe that Theorem 1 implies that the regret of the SOAR algorithm can be written as the average of the regret of a sequence of hindsight problems, i.e.,*  
 $\text{REG}_n(\text{SOAR}) = U_\infty - U_n(\text{SOAR}) = \frac{1}{n} \sum_{k=1}^n (U_\infty - U_k^{\text{H}}) \triangleq \frac{1}{n} \sum_{k=1}^n \text{REG}_k(\text{H-OPT})$ , *where H-OPT stands for the hindsight optimal algorithm.*

Theorem 1 reveals the connection between the performance of SOAR and the convergence

behavior of the sequence of hindsight optimum values. Rather surprisingly, Theorem 1 implies that SOAR achieves vanishing regret under no assumption other than the boundedness of  $\varphi$ , as we summarize in the next Corollary.

**Corollary 1** *Under the same assumption as in Theorem 1, SOAR achieves vanishing regret:  $\lim_{n \rightarrow \infty} \text{REG}_n(\text{SOAR}) = 0$ .*

The proof follows immediately from Remark 1 and the monotonicity of  $U_k^H$ , which we delegate to Appendix A.3. When the matching instance satisfies additional conditions, Theorem 1 further implies the regret scaling of SOAR. In particular, we expect the regret of SOAR to converge at the same rate as the regret of H-OPT as  $n$  grows in most spatial matching settings, as the former is the Cesàro sum of the later via Remark 1. To formalize this intuition, we first introduce the following regular-scaling property of an offline bipartite matching instance specified by primitives  $P, Q$ , and  $\varphi$ .

**Definition 1 (Regular and polynomial regret scaling)** *A matching instance  $(P, Q, \varphi)$  is said to scale regularly with parameter  $\beta > 0$  if, for any  $0 < \epsilon < \beta$ , we have  $\limsup_{n \rightarrow \infty} n^{\beta - \epsilon} \cdot (U_\infty - U_n^H) = 0$  and  $\liminf_{n \rightarrow \infty} n^{\beta + \epsilon} \cdot (U_\infty - U_n^H) = \infty$ . If in addition  $\lim_{n \rightarrow \infty} n^\beta \cdot (U_\infty - U_n^H) = l_0$  for some constant  $l_0$ , then we say the matching instance  $(P, Q, \varphi)$  scales polynomially with parameter  $\beta$ .*

No policy can do better than the hindsight optimal policy and, in particular, we have  $\text{REG}_n(\pi) \geq \text{REG}_n(\text{H-OPT})$  for any  $\pi$ . Under regular or polynomial scaling, the following corollary of Theorem 1 tells us that  $\text{REG}_n(\text{SOAR})$  scales like  $\text{REG}_n(\text{H-OPT})$ .

**Corollary 2** *For a matching instance that scales regularly with parameter  $\beta > 0$ , we have that, for any  $\epsilon > 0$ ,  $\limsup_{n \rightarrow \infty} n^{\beta - \epsilon} \text{REG}_n(\text{SOAR}) = 0$  and  $\liminf_{n \rightarrow \infty} n^{\beta + \epsilon} \text{REG}_n(\text{SOAR}) = \infty$ , and*

$$\text{REG}_n(\text{SOAR}) \leq n^\epsilon \cdot \text{REG}_n(\text{H-OPT}) \text{ for } n \text{ sufficiently large.}$$

*If, in addition, the matching instance scales polynomially with parameter  $\beta$ , then there exists a*

constant  $l_1 = l_1(P, Q, \varphi)$  independent of  $n$ , such that for all  $n \in \mathbb{N}$

$$\text{REG}_n(\text{SOAR}) \leq \begin{cases} l_1 \text{REG}_n(\text{H-OPT}), & \text{if } \beta \in (0, 1), \\ l_1 \log n \cdot \text{REG}_n(\text{H-OPT}), & \text{if } \beta = 1. \end{cases}$$

Furthermore, in this case, the regret scaling of **SOAR**, which is also the optimal regret scaling (up to at most a logarithmic factor), can be tightly characterized:

$$\text{REG}_n(\text{SOAR}) = \begin{cases} \Theta(n^{-\beta}), & \text{if } \beta \in (0, 1), \\ \Theta(n^{-1} \log n), & \text{if } \beta = 1. \end{cases}$$

This corollary shows that **SOAR** enjoys a universal guarantee of near-optimal regret scaling when the matching instance scales regularly. We note that regular (or polynomial) scaling seems to be a relatively mild requirement, which is satisfied by a number of matching instances of practical interest. For example, when both demand and supply units are uniformly distributed on  $[0, 1]^d$  and the cost function is  $-\|X - Y\|^p$ ,  $U_\infty = 0$ , the hindsight optimum scales regularly for all  $d, p \geq 2$ , with  $\beta = p/d$  [52]. Then it follows immediately from Corollary 2 that in these examples, **SOAR** provides a tight regret scaling for all  $d, p \geq 2$ . Refer to Appendix A.5 for details. Indeed, we expect that matching instances satisfying certain basic conditions (e.g., continuity or boundedness of quality function  $\varphi$ ) should all scale regularly or even polynomially. Determining primitive conditions on  $P, Q, \varphi$  under which the associated matching instances scale regularly or polynomially is in general a mathematically intriguing question, and we defer further investigation to future research. We defer the proof of the corollary to Appendix A.4.

To determine the precise value of  $\beta$  for a particular matching instance  $(P, Q, \varphi)$  can be quite challenging (see [52]). In many cases, we are often only able to establish bounds on the hindsight regret  $\text{REG}_n(\text{H-OPT})$ . Indeed,  $\text{REG}_n(\text{H-OPT})$  emerges as the objective of an extensively studied problem: *Empirical optimal transport* (see [53], Appendix A.6.1). Its rich literature provides upper bounds on  $\text{REG}_n(\text{H-OPT})$  for matching instances belonging to various particular classes,

e.g.  $P, Q$  satisfying certain smoothness conditions and/or  $\varphi(X, Y) = -\|X - Y\|^p$ . Leveraging the analytical framework set up in this section, we immediately get upper bounds on the regret scaling of SOAR for these classes of matching instances. Such upper bounds are tight if they are matched by an instance in these classes. In the next section, we take the above approach to establish the near-optimal regret scaling of SOAR for some interesting classes of matching instances.

#### 1.4 Near-optimal Regret Scaling of SOAR for the $-\|X - Y\|^p$ and $\langle X, Y \rangle$ Quality Functions

In this section, we leverage our analytical framework (Theorem 1) to establish the near-optimal regret scaling of SOAR for a general quality function  $\varphi_p(X, Y) = -\|X - Y\|^p$  for  $p \geq 1$ . Two important cases of this general quality function are  $p = 1$  and  $p = 2$ . Our matching problem for the case of  $p = 1$  (euclidean distance cost) corresponds to the setting studied in the semi-dynamic model of [10] assuming uniform supply and demand distributions. The Euclidean distance cost has been widely studied in the context of ride hailing platforms [14, 10, 54]. In terms of regret scaling, the quality function  $\varphi_p(X, Y) = -\|X - Y\|^p$  for the case of  $p = 2$  is equivalent to the dot product quality function  $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$  (see Lemma 10), which is a practically motivated quality function inspired by recommender systems [55, 56], a close cousin of our matching problem. To obtain our results, we leverage the fact that the offline version of our matching problem for the quality function  $\varphi_p(X, Y) = -\|X - Y\|^p$  is intimately related to the problem of quantifying the empirical Wasserstein- $p$  distance between two measures, which has been extensively studied in the literature ([57, 58, 59, 60]). The meta performance analysis (Theorem 1) allows us to directly infer from these results the regret scaling of SOAR in the corresponding online setting. We obtain two sets of regret scalings achieved by SOAR: one set allows for general distributions  $P$  and  $Q$  (Theorem 2 and Corollary 3), and the other set requires  $P$  and  $Q$  to be sufficiently smooth (Proposition 2 and Theorem 3). For the latter, we observe sharper regret scaling in low dimensions ( $d \leq 3$ ). We construct hard instances to demonstrate the tightness of these regret scalings.

### 1.4.1 Performance Guarantees for $-\|X - Y\|^p$ Quality Functions

We begin by providing near-optimal regret guarantees for **SOAR** for the  $\varphi_p(X, Y) = -\|X - Y\|^p$  quality functions under very general assumptions on the demand and supply distributions. The proof of Theorem 2 is deferred to Appendix A.7.

**Theorem 2** *Suppose  $P$  and  $Q$  are supported on bounded sets. Under the quality function  $\varphi_p(X, Y) = -\|X - Y\|^p$  for  $p \geq 1$ , there exists a universal constant  $C := C(P, Q, d, p) < \infty$  such that the regret of **SOAR** is bounded above as*

$$\text{REG}_n(\text{SOAR}) \leq \begin{cases} Cn^{-\frac{1}{2}}, & d < 2(p \wedge 2), \\ Cn^{-\frac{1}{2}} \log n, & d = 2(p \wedge 2), \\ Cn^{-\frac{p \wedge 2}{d}}, & d > 2(p \wedge 2). \end{cases}$$

Furthermore, the aforementioned regret scaling is nearly the best possible, formalized as follows. For each  $d \geq 1$ , there exists a pair of distributions  $P$  and  $Q$  (supported on bounded sets) such that there exists a constant  $c := c(P, Q, d, p) > 0$  and the optimal regret scaling is bounded below as

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi) \geq \begin{cases} cn^{-\frac{1}{2}}, & d < 2(p \wedge 2), \\ cn^{-\frac{1}{2}}, & d = 2(p \wedge 2), \\ cn^{-\frac{p \wedge 2}{d}}, & d > 2(p \wedge 2). \end{cases}$$

In contrast to myopic policies like **Greedy** that incurs non-vanishing regret (cf. Proposition 1), our forward-looking simulation-based policy **SOAR** achieves vanishing regret. Moreover, the upper bounds when viewed in conjunction with their corresponding lower bounds establish the near-optimality of **SOAR**. We observe that the regret increases with dimension  $d$ . This is due to an increase in the intrinsic hardness of matching, cf. [10]. Indeed, the available supply units become sparser in higher dimensional spaces. Thus finding supply units that are compatible with the demand units along all dimensions becomes harder, resulting in larger regret.

**Remark 2 (Dependence of constants)** *The universal constant  $C$  in the upper bound in Theorem 2 is exponential in the dimension  $d$ . In contrast, the constant  $c$  in the lower bound scales polynomially in the dimension  $d$ . It is unclear if the exponential dependence on  $d$  of the constant  $C$  is tight or merely an artifact of the analysis and we leave the investigation of this question and possibly improving the dependence of the constant  $C$  on dimension  $d$  for future research.*

Under more stringent conditions on  $P$  and  $Q$ , sharper regret scalings are achievable. We state such a result for the special case of  $P = Q = \text{Uniform}([0, 1]^d)$ .

**Proposition 2** *Suppose  $P = Q = \text{Uniform}([0, 1]^d)$ . Under the quality functions  $\varphi_p(X, Y) = -\|X - Y\|^p$  for  $p \geq 1$ , there exists a universal constant  $C := C(P, Q, d, p) < \infty$  such that*

$$\text{REG}_n(\text{SOAR}) \leq \begin{cases} Cn^{-(\frac{p}{2} \wedge 1)} \mathbb{1}\{p \neq 2\} + C(n^{-1} \log n) \mathbb{1}\{p = 2\}, & d = 1, \\ C(n^{-1} \log n)^{\frac{p}{2}} \mathbb{1}\{p < 2\} + C(n^{-1} (\log n)^2) \mathbb{1}\{p = 2\} + Cn^{-1} \mathbb{1}\{p > 2\}, & d = 2, \\ Cn^{-(\frac{p}{d} \wedge 1)} \mathbb{1}\{p \neq d\} + C(n^{-1} \log n) \mathbb{1}\{p = d\}, & d \geq 3. \end{cases}$$

Furthermore, the aforementioned regret scaling can not be improved in general. For each  $d \geq 1$ , there exists a constant  $c := c(d) > 0$  and the optimal regret scaling is bounded below as

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi) \geq \begin{cases} cn^{-(\frac{p}{2} \wedge 1)} \mathbb{1}\{p \neq 2\} + cn^{-1} \mathbb{1}\{p = 2\}, & d = 1, \\ c(n^{-1} \log n)^{\frac{p}{2}} \mathbb{1}\{p < 2\} + c(n^{-1} \log n) \mathbb{1}\{p = 2\} + cn^{-1} \mathbb{1}\{p > 2\}, & d = 2, \\ cn^{-(\frac{p}{d} \wedge 1)} \mathbb{1}\{p \neq d\} + c(n^{-1} \log n) \mathbb{1}\{p = d\}, & d \geq 3. \end{cases}$$

We defer the proof of Proposition 2 to Appendix A.8. We note that in comparison with Theorem 2, the performance of SOAR improves for  $P = Q = \text{Uniform}([0, 1]^d)$  in various parameter regimes. For example, in Theorem 2 where  $P$  and  $Q$  can be general distributions supported on bounded sets, for  $p \in (2, d]$  and  $d \geq 4$ , the regret scales as  $\Theta(n^{-2/d})$ . However, in Proposition 2, the regret scaling improves to  $\tilde{\Theta}(n^{-p/d})$ . There are two main drivers for the sharper regret scaling: (i) smoothness of the demand and supply distribution and (ii) both the demand and supply distribu-

tions being identical. In light of the hard instance for Theorem 2 that consists of discrete distributions (cf. proof of Theorem 2 in Appendix A.7), the smoothness of distribution  $\text{Uniform}([0, 1]^d)$  in Proposition 2 seems necessary for the sharper regret scalings. Later in Subsection 1.4.2, we present a similar regret guarantee in the case of  $p = 2$  (equivalently, the case of dot-product quality function) (cf. Theorem 3), for a more general class of  $P$  and  $Q$  that allow for possibly non-uniform and unequal distributions, assuming some proper smoothness conditions. However, such an extension is not immediate for general  $p \neq 2$  due to lack of regularity of the optimal transport map ([61]), and we defer a more in-depth exploration of the assumptions that may yield sharper regret scaling for general quality function to future research.

*Open Problem in [10].* The case of  $p = 1$  (euclidean distance cost) has been extensively studied for all  $d \geq 1$  in [10]. [10] develops a greedy-like algorithm called Hierarchical Greedy and shows that Hierarchical Greedy achieves near-optimal regret scaling for all  $d \geq 1$ . However, this near optimality does not hold for general  $p > 1$ : for a fixed  $d \geq 2$  and  $p \in (d/2, d]$ , the regret of Hierarchical Greedy scales as  $\Theta(n^{-1/2})$  whereas a sharper regret scaling of  $\tilde{\Theta}(n^{-p/d})$  is achievable via SOAR and up to logarithmic factors, this regret scaling is tight. In fact, [10] leaves it as an open problem to close the gap between the regret scaling achieved via Hierarchical Greedy and optimal regret scaling of  $\tilde{\Omega}(n^{-p/d})$  (assuming  $d \geq 2, p \leq d$ ) for uniform demand and supply distributions and  $\varphi_p(X, Y) = -\|X - Y\|^p$ . As a consequence of Proposition 2, we resolve this open problem.

#### 1.4.2 Performance Guarantees for the $\langle X, Y \rangle$ Quality Function

We first observe that in terms of regret, the dot-product quality function  $\langle X, Y \rangle$  is equivalent to  $\varphi_2(X, Y) = -\|X - Y\|^2$  (cf. Lemma 10). Therefore, the regret scaling of SOAR for the dot-product quality function under the assumptions of the demand and supply distributions being supported on bounded sets follows immediately from Theorem 2.

**Corollary 3** *Suppose  $P$  and  $Q$  are supported on bounded sets. The regret scalings for the dot product function  $\varphi(X, Y) = \langle X, Y \rangle$  correspond to the regret scalings demonstrated in Theorem 2*

with  $p = 2$ .

**Remark 3 (Near-optimal regret scaling with scarce supply and rejection cost)** *Recall that we modelled scarce supply and rejection cost in Section 5.2 by considering an augmented demand and supply distribution  $P'$  and  $Q'$  and a modified quality function. For the particular quality function  $\varphi(X, Y) = \langle X, Y \rangle$ , its corresponding quality function that incorporates the rejection cost can be reformulated so as to retain the dot-product form, albeit in a different, augmented feature space. The key idea is to view the rejection cost as an additional dimension for the demand unit's weight vector. As before, let  $c(x)$  denote the cost of rejection for a demand unit  $x \in \mathcal{X}$  that is measurable and bounded. Let  $\mathbf{dum} = (\mathbf{0}_{d \times 1}, 1)$  and define the following  $d + 1$  dimensional spaces*

$$\mathcal{X}' = \{(x, -c(x)) : x \in \mathcal{X}\} \subseteq \mathbb{R}^{d+1}, \quad \mathcal{Y}' = \{(y, 0) : y \in \mathcal{Y}\} \cup \mathbf{dum} \subseteq \mathbb{R}^{d+1}.$$

Then for any  $x' \in \mathcal{X}'$  and  $y' \in \mathcal{Y}'$ , we have that  $\varphi'(x', y') = \langle x, y \rangle \mathbb{1}\{y \neq \mathbf{dum}\} - c(x) \mathbb{1}\{y = \mathbf{dum}\} = \langle x', y' \rangle$ . Let  $P'$  denote a distribution supported on  $\mathcal{X}'$  and  $Q'$  denote a distribution supported on  $\mathcal{Y}'$ . As previously discussed,  $Q'$  is a mixture distribution of  $Q$  and  $Q^{\mathbf{dum}}$ . By modelling the virtual supply unit  $\mathbf{dum}$  in the distribution  $Q'$  and the rejection cost in distribution  $P'$ , we are effectively back to the balanced setting with  $n$  supply units  $Y'_1, Y'_2, \dots, Y'_n$  independently sampled from  $Q'$  and  $n$  demand units  $X'_1, X'_2, \dots, X'_n$  independently sampled from  $P'$ , associated with the dot-product quality function. Furthermore, we note that  $P'$  and  $Q'$  will be supported on bounded sets as long as  $P$  and  $Q$  are supported on bounded sets and  $c$  is a measurable bounded function. Hence from Corollary 3, even under scarce supply, we are able to achieve near-optimal regret scaling for the quality function  $\varphi(X, Y) = \langle X, Y \rangle$ . Note that the benchmark in the case of scarce supply is  $U_\infty(P', Q')$ .

**Remark 4 (Extension to polynomial kernel quality functions)** *Our regret bound on SOAR for the dot product quality function implies regret bounds for a broader class of quality functions which we refer to as the polynomial kernel quality functions. For some  $q \in \mathbb{N}$ , we refer to  $\varphi_{\text{ker}}^q(X, Y) = \langle X, Y \rangle^q$  as the polynomial kernel quality function. Using Corollary 3, one can easily establish*

vanishing regret for SOAR under this more general class of polynomial kernel quality functions  $\varphi_{\ker}^q(X, Y)$  (see Corollary 11). The key idea is the well-known result that for each choice of  $q \in \mathbb{N}$ , there exists a mapping  $\phi_q : \mathbb{R}^d \rightarrow \mathbb{R}^{d_q}$  where  $d_q = \binom{d+q-1}{q}$  such that  $\varphi_{\ker}^q(X, Y) = \langle X, Y \rangle^q = \langle \phi_q(X), \phi_q(Y) \rangle$  [62]. This dot product representation allows us to invoke Corollary 3 and establish vanishing regret. Furthermore, the vanishing regret guarantee of SOAR can be extended to a conic combination (weighted sum with non-negative coefficients) of polynomial kernel quality functions corresponding to different values of  $q \in \mathbb{N}$ , i.e.,  $\varphi_{\ker}(X, Y) = \sum_{q=0}^m a_q \varphi_{\ker}^q(X, Y) = \sum_{q=0}^m a_q \langle X, Y \rangle^q$  (see Corollary 11). For further details, refer to Appendix A.9.

In Proposition 2, we showed that sharper regret scalings are achievable by SOAR when the demand and supply distributions are uniform. Under the dot-product quality function, we can further generalize those results to settings where demand and supply distributions are smooth (uniform-like) and possibly distinct and under a key curvature condition on the *Brenier potential*. The Brenier potential ([63]) is a convex function  $\psi_0$  whose gradient uniquely defines the optimal transport map  $\mathcal{T}_{P \rightarrow Q}$  from  $P$  to  $Q$  under the quadratic cost (equivalently, the dot-product quality function), for  $P$  and  $Q$  satisfying some smoothness assumptions. We refer the interested readers to [64] for a detailed introduction of the Brenier potential and the related background knowledge of optimal transport (a brief discussion is provided in Appendix A.6).

**Assumption 1 (Curvature condition)** *The Brenier potential  $\psi_0$  is a closed, convex, function such that  $\psi_0 \in C^2([0, 1]^d)$  and  $(1/\lambda)\mathbf{I}_{d \times d} \preceq \nabla^2 \psi_0(x) \preceq \lambda \mathbf{I}_{d \times d}$  for all  $x \in [0, 1]^d$  and for some  $\lambda \geq 1$ .*

The curvature condition on the the Brenier potential in Assumption 1 has been borrowed from [60] and it enables us to prove sharper regret scaling under smoothness assumptions of the demand and supply distributions. This sharper regret scaling is formalized in Theorem 3 presented below.

**Theorem 3** *Suppose  $P$  and  $Q$  are absolutely continuous distributions on  $[0, 1]^d$  with densities  $p$  and  $q$ . Assume that there exists  $\gamma_p \geq 1$  and  $\gamma_q \geq 1$  such that  $\gamma_p^{-1} \leq p \leq \gamma_p$  and  $\gamma_q^{-1} \leq q \leq \gamma_q$  over  $[0, 1]^d$  and further assume the curvature condition in Assumption 1. Under the quality function*

$\varphi(X, Y) = \langle X, Y \rangle$  there exists a universal constant  $C := C(P, Q, d) < \infty$  such that the regret of SOAR is bounded above as

$$\text{REG}_n(\text{SOAR}) \leq \begin{cases} Cn^{-1} \log n, & d = 1, \\ Cn^{-1} (\log n)^3, & d = 2, \\ Cn^{-\frac{2}{d}}, & d \geq 3. \end{cases}$$

Furthermore, the aforementioned regret scaling can not be improved in general. The lower bound on the optimal regret scaling with quality function  $\varphi(X, Y) = \langle X, Y \rangle$  follows from Proposition 2 for the case of  $p = 2$ .

We defer the proof of Theorem 3 to Appendix A.10. Comparing Corollary 3 with Theorem 3 for  $\varphi(X, Y) = \langle X, Y \rangle$ , we observe that for  $d \leq 3$ , Theorem 3 provides a sharper regret scaling: contrast the regret scaling of  $\Theta(n^{-1/2})$  for  $d \leq 3$  and  $p = 2$  in Corollary 3 with  $\tilde{\Theta}(n^{-1})$  for  $d = 1, 2$  and  $\Theta(n^{-2/3})$  for  $d = 3$  in Theorem 3. The special structure of the dot-product quality function  $\varphi(X, Y) = \langle X, Y \rangle$  (a special case of  $\varphi_p(X, Y) = -\|X - Y\|^p$  for  $p = 2$ ) enables us to derive sharper regret scalings under some smoothness assumptions of the demand  $P$  and supply  $Q$  distributions and the curvature condition in Assumption 1.

## 1.5 Numerics

In this section, we describe our numerical simulations for different dimensions  $d$  and different sets of demand and supply distributions. We focus on the quality function  $\varphi(X, Y) = -\|X - Y\|^2$  unless mentioned otherwise, where we recall the equivalence between this quality function and  $\varphi(X, Y) = \langle X, Y \rangle$  in terms of regret. We consider different demand and supply distributions and evaluate the performance of different algorithms as listed in Table 1.2. In essence, our numerical experiments exemplify our theoretical guarantees that our proposed algorithm SOAR achieves vanishing regret at near-optimal rate in all the settings we consider whereas myopic policies like Greedy or OT + Greedy (a smarter variant of Greedy) either suffer from non-vanishing regret or

the rate is sub-optimal. In the following subsections, we describe the settings considered in detail and present the regret scaling for different algorithms. Note that all the plots are presented in the log – log scale.

Table 1.2: Summary of algorithms considered for numerical simulations

Algorithm	Requires OT Map	Description
SOAR	No	Algorithm 1
Greedy	No	Match $X_t$ to nearest supply
OT + Greedy	Yes	Transport $X_t$ to $Q$ and match to nearest supply
Hierarchical Greedy	N/A	Algorithm 1 in [10]

### 1.5.1 Setting (I) $P = Q = \text{Uniform}([0, 1]^d)$

We consider the case of demand and supply distributions being equal and in particular, being  $\text{Uniform}([0, 1]^d)$ . For this setting, we present the results for  $d \in \{1, 2, 3\}$  in Figure 1.1. We compare the performance of the Hierarchical Greedy ([10]), Greedy, and SOAR algorithms. In Figure 1.1, we observe that as the number of supply units  $n$  increases, SOAR performs better than Greedy and Hierarchical Greedy. Observe the slopes corresponding to the different algorithms. Both Hierarchical Greedy and Greedy algorithms have a slope close to  $-0.5$  which corresponds to the regret of these algorithms scaling as  $O(n^{-1/2})$ . For SOAR, the slope for different values of  $d$  closely matches the theoretical regret guarantees of  $O(\log n/n)$  for  $d = 1$  (slope is  $-0.82$ ),  $O(\log^3 n/n)$  for  $d = 2$  (slope is  $-0.78$ ) and  $O(n^{-2/3})$  for  $d = 3$  (slope is  $-0.64$ ).

### 1.5.2 Setting (II) $P = \text{Uniform}([0, 1]^d), Q = \text{Uniform}([0, 2]^d)$

We consider an example where the demand and supply distributions are unequal and the optimal transport map is fairly easy to compute i.e  $\mathcal{T}_{P \rightarrow Q}(X) = 2X$ . This enables us to implement the OT + Greedy algorithm, which is a smarter variant of Greedy. The OT + Greedy algorithm

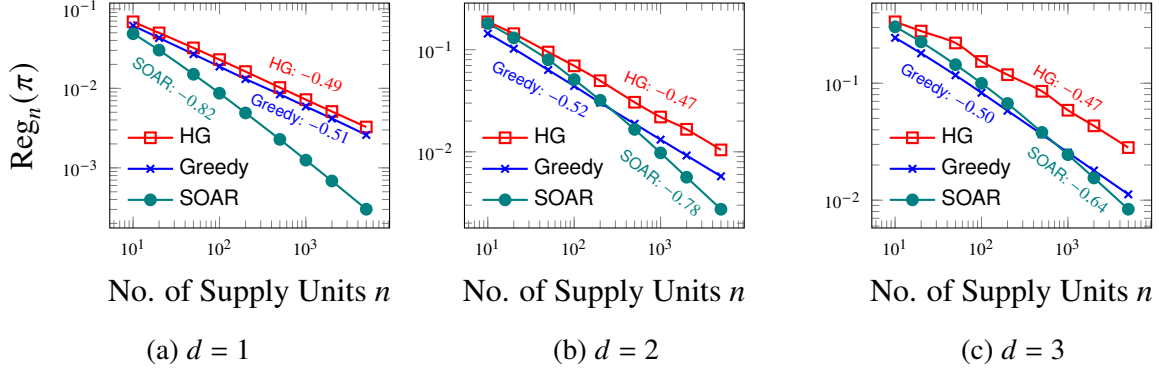


Figure 1.1: Comparing the performance of Hierarchical Greedy (HG), Greedy and SOAR for  $P = Q = \text{Uniform}([0, 1]^d)$ .

is as follows: upon observing a demand request  $X_t$ , we first transport the demand request  $X_t$  using the optimal transport map  $\mathcal{T}_{P \rightarrow Q}$  and then *greedily* match the transport demand unit  $\mathcal{T}_{P \rightarrow Q}(X_t)$  to its nearest existing supply unit. By computing the transport map of the demand units, the OT + Greedy algorithm reduces to a dynamic greedy matching between random points of the same distribution. Moreover, instead of directly implementing a greedy matching between the demand unit and existing supply units, by utilizing the optimal transport map  $\mathcal{T}_{P \rightarrow Q}$ , we are able to make the Greedy algorithm forward looking. For this setting, we present the results for the case of  $d \in \{1, 2, 3\}$  in Figure 1.2. Note that this setting is equivalent to the previous setting upto the optimal transport. In this setting, we observe that the Greedy algorithm suffers from non-vanishing regret and this is in line with Proposition 1. Both the smarter variant of Greedy, which we dub as OT + Greedy, and SOAR are able to achieve vanishing regret. However, we note that OT + Greedy requires nontrivial knowledge of the underlying model, namely the optimal transport map between  $P$  and  $Q$ , which quickly becomes a computation burden in more complex settings. Moreover, observe the slopes of the curve corresponding to OT + Greedy and SOAR algorithm. The slope of the curves corresponding to SOAR are steeper compared to the ones corresponding to OT + Greedy for  $d \in \{1, 2, 3\}$  which implies that SOAR achieves a sharper regret scaling in comparison to the OT + Greedy policy.

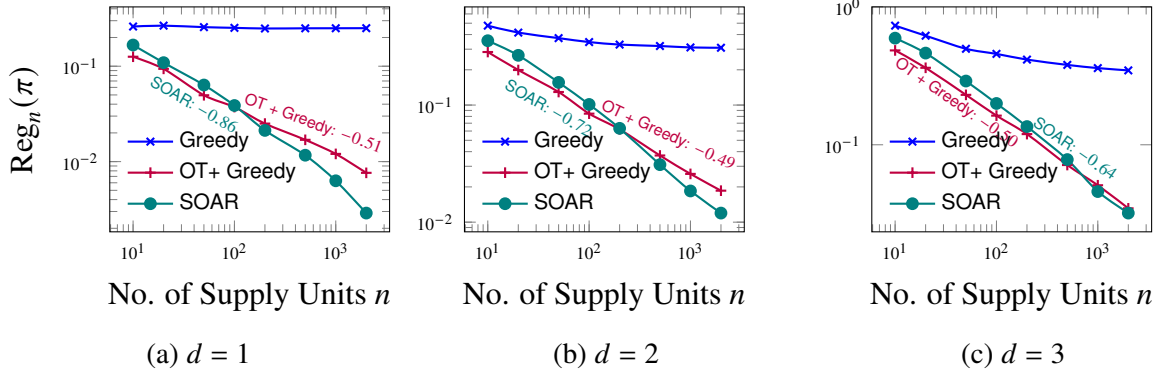


Figure 1.2: Comparing the performance of Greedy, OT + Greedy and SOAR for  $P = \text{Uniform}([0, 1/2]^d)$  and  $Q = \text{Uniform}([0, 1]^d)$ .

### 1.5.3 Setting (III) $P = \text{TruncNorm}(\mu, \Sigma)$ , $Q = \text{Uniform}([0, 1]^d)$

We consider an example where the demand and supply distributions are unequal however it is non-trivial to compute the optimal transport map. In particular, we assume that the demand distribution is a truncated normal with mean  $\mu = (1/2) \times \mathbb{1}_{d \times 1}$  and covariance  $\Sigma = 0.1 \times \mathbb{I}_{d \times d}$  and the supply distribution is  $\text{Uniform}([0, 1]^d)$ . Since the optimal transport map is non-trivial to compute, we approximate the value of the fluid optimum for each  $d$  via simulation. For this setting, we present the results for the case of  $d \in \{1, 2, 3\}$  in Figure 1.3 and compare the Greedy algorithm and the SOAR algorithm. As before, we observe that the SOAR algorithm outperforms the Greedy algorithm.

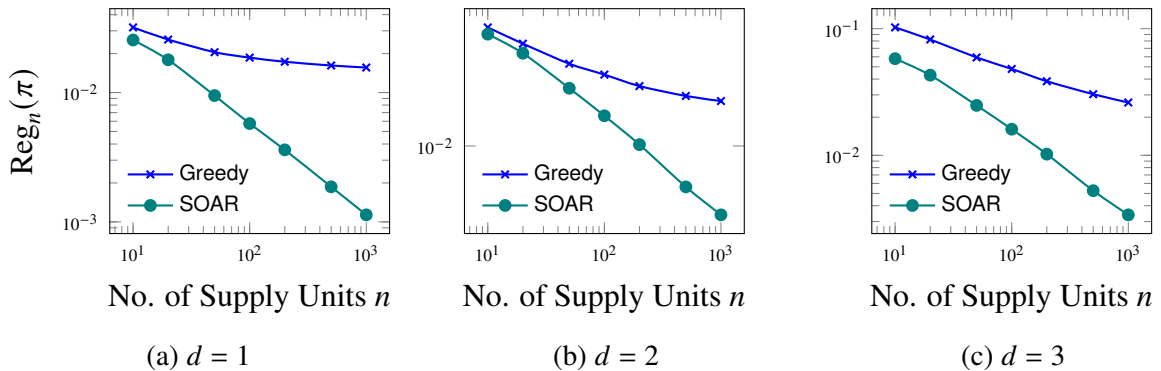


Figure 1.3: Comparing the performance of Greedy and SOAR for  $P = \text{TruncNorm}(\mu, \Sigma)$  and  $Q = \text{Uniform}([0, 1]^d)$ .

1.5.4 Setting (IV)  $P = \text{Uniform}([0, 1]^d)$ ,  $Q = \text{TruncNorm}(\mu_{d-2 \times 1}, \Sigma_{d-2 \times d-2}) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$

We consider the case where demand and supply distributions are unequal, the optimal transport is non-trivial to compute and moreover, the supply distribution is not smooth unlike the previously consider settings. We focus on the case of  $d \geq 3$ . The supply distribution is  $Q = \text{TruncNorm}\left(\frac{1}{2} \times \mathbb{1}_{d-2 \times 1}, 0.1 \times \mathbb{I}_{d-2 \times d-2}\right) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$ . Since the optimal transport map is non-trivial to compute, we approximate the fluid optimum value via simulation. For this setting, we present the results for the case of  $d \in \{3, 4, 5\}$  in Figure 1.4 and compare the Greedy and SOAR algorithms. We observe that as  $n$  increases, the performance of SOAR dominates the performance of Greedy algorithm.

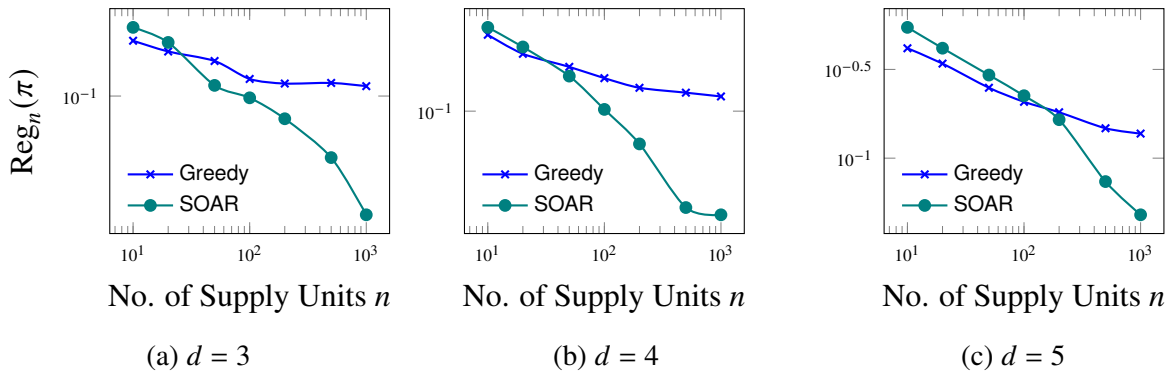


Figure 1.4: Comparing the performance of Greedy and SOAR for  $P = \text{Uniform}([0, 1]^d)$  and  $Q = \text{TruncNorm}(\mu, \Sigma) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$ .

## 1.6 Conclusion and Future Research

In this work, we study a dynamic two-sided matching problem as the market thickness  $n$  scales, and characterize the optimal regret scaling for all dimensions  $d$ . We develop a principled simulation-based approach dubbed Simulate-Optimize-Assign-Repeat (SOAR) and demonstrate that this forward-looking policy is vastly superior to myopic policies like Greedy. En-route we develop a novel framework for regret analysis where we provide a simple formula connecting the performance of SOAR with a sequence of hindsight optimum values. As a corollary of our

techniques, we also resolve one of the open problems in [10].

Our algorithm **SOAR** and results crucially rely on knowledge of the horizon  $n$ . Given the recent burgeoning interest in the study of online resource allocation problems in the presence of horizon uncertainty [65, 16, 66, 67], an interesting follow-up would be to design and analyze near-optimal algorithms for such uncertain scenarios. One approach to modelling horizon uncertainty is to assume distributional knowledge of the horizon [as done in 67, 16], where the horizon length (total demand units) is modelled as a random variable  $N$  and this is known to the platform. Assume that  $n = \mathbb{E}[N]$ , then in assigning  $n$  supply units to the  $N$  arriving demand units, **SOAR** can be implemented by fixing the number of demand units as if it was  $n$ . If in addition,  $N$  is well concentrated, i.e.,  $\text{var}(N) = o(n^2)$  (e.g.,  $N$  is a Binomial or Poisson random variable), **SOAR** achieves vanishing regret where the rate may depend on the variance of  $N$ . On the other hand, if the variance is large, i.e.,  $\text{var}(N) = \Omega(n^2)$ , then **SOAR** suffers from non-vanishing regret and this observation is in line with other resource allocation works studying horizon uncertainty [16, 67]. The large variance case is quite challenging and developing near-optimal algorithms would require novel approaches and we leave this line of research for future work.

## **Chapter 2: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances**

*Based on the paper [68] co-authored with Omar Besbes and Yash Kanoria.*

### **2.1 Introduction**

Online resource allocation provides a comprehensive framework for scenarios that involve allocating finite resources to requests arriving over time, with the objective of maximizing the overall reward. This model encompasses several well-studied problems, such as the multisecretary problem [69, 70], network revenue management [71, 36, 37], and order fulfillment [72].

Prior work mainly explores these problems under one of two distributional assumptions on the request types: (i) atomic distributions supported on a few points [36, 37] and (ii) non-atomic distributions with contiguous support [73, 70]. Under these cases, impressive constant and logarithmic regret guarantees have been established, where regret is defined as the expected difference between the total reward under the optimal hindsight policy and the total reward gathered under an online policy.

However, for many important applications, neither of these two assumptions adequately capture reality. For instance, consider the order fulfillment problem encountered by e-commerce platforms like Amazon or Walmart. This is an online matching problem with spatially distributed demand (different zip codes or counties) with product inventory housed in various warehouses scattered across a geographic area. The fulfillment team aims to minimize cumulative shipping costs by dynamically matching each demand to a warehouse which has the item available. Warehouses have limited inventory, and decisions must be made in real-time. This problem can be framed within the online resource allocation problem paradigm. Yet, the aforementioned assumptions

made in the prior literature do not capture key features of this setting: (i) the number of demand locations (types) is large (for instance, there are over 40,000 zip codes in the United States), and (ii) these demand locations are spatially clustered with gaps (regions with no demand), a natural characteristic of geographical landscapes such as rivers, mountains, deserts, etc. Hence, atomic distributions with a low number of types or non-atomic distributions with contiguous support fail to capture the salient features of such a problem. Aside from modeling concerns, the near-optimal algorithms developed for each of the two classes of distributions mentioned above are tailored to that particular class of distributions.

The above motivation leads us to the following research questions: (i) *What (request type) distribution features drive achievable performance, and how does regret scale as a function of the underlying distribution?* (ii) *What algorithmic principles allow one to achieve optimal regret scaling?* (iii) *Is there a unifying near-optimal algorithm that is agnostic to the underlying distribution's features?*

To isolate and examine key performance drivers, we will initially focus on one of the simplest online resource allocation problems: the multiselection problem, which is a special case of both the network revenue management problem as well as the online matching (order fulfillment) problem (we refer to Appendix B.7 for a more extensive discussion on the latter connection). In the multiselection problem, a decision-maker (DM) with a budget to hire  $B$  secretaries is presented with a series of  $T$  independent values representing candidate abilities. The DM must make irrevocable “accept” (i.e., hire) or “reject” decisions on the fly, aiming to maximize the (expected) sum of the chosen candidates’ abilities.

We make three main contributions. The first two are in the context of the multiselection problem: fundamental lower bounds on regret, and an algorithmic principle to achieve the optimal regret scaling. Our third contribution is a unifying and practical algorithm for achieving near optimal regret performance in general resource allocation problems. We now elaborate on these contributions.

(i) *Drivers of regret*: In the context of the multiselection problem, we identify a novel funda-

mental driver of regret which is characterized by a parameter  $\beta$ , which quantifies the mass accumulation of types around gaps (interval with zero probability mass). Using this parameter  $\beta$  we characterize a broad class of distributions with gaps, which we refer to as  $(\beta, \varepsilon_0, \delta)$ -clustered distributions (cf. Definition 2). The class of  $(\beta, \varepsilon_0, \delta)$ -clustered distributions is a superset of the class of discrete distributions [69], and the class of non-atomic distributions with continuous support over  $[0, 1]$  and density uniformly bounded away from zero [70]. We establish a universal lower bound (for any policy) on the growth rate of the regret as a function of the parameter  $\beta$  which quantifies how mass accumulates around gaps. In particular, we establish that any policy must incur  $\Omega(T^{\frac{1}{2} - \frac{1}{2(1+\beta)}})$  regret in the worst-case (cf. Theorem 4) for a  $(\beta, \varepsilon_0, \delta)$ -clustered distribution. This is in stark contrast to prior results which prove regret scaling of  $\Theta(1)$  [69] for the case of distributions with a few discrete types and  $\Theta(\log T)$  [70] for a special class of non-atomic distributions. We also show that our lower bound on the regret scaling is achievable up to polylogarithmic factors. To the best of our knowledge, ours is the first result of its kind; notably the regret scaling we establish is polynomial in  $T$  for  $\beta > 0$  and an entire spectrum of regret scalings are possible. As  $\beta$  increases, so does the exponent  $\frac{1}{2} - \frac{1}{2(1+\beta)}$  (from 0 to 1/2), characterizing the “hardness” of the problem instance.

- (ii) Algorithmic Principle: It turns out the workhorse certainty equivalent (CE) policy is insufficient to deal with general type distributions which have gaps in the support, already in the case of the multisecretary problem. For such distributions, we introduce a new algorithmic principle we call *Conservativeness with respect to gaps* (CwG); which makes a crucial modification to the CE policy. The idea is that if at any time the CE threshold is close to the boundary of a gap, CwG instead uses the gap as the acceptance threshold to avoid incurring large regret in the future. We establish that this enables the policy to mitigate the risk of incurring large regret (in the event that the threshold for the hindsight optimal falls on the opposite side of that gap). We use this principle to design a near-optimal algorithm, dubbed CwG, for the  $(\beta, \varepsilon_0, \delta)$ -clustered distributions. Its worst-case regret scales as  $\tilde{O}(T^{\frac{1}{2} - \frac{1}{2(\beta+1)}})$ ,

matching the scaling of the lower bound in  $T$  up to polylogarithmic terms (cf. Theorem 5). For the case of a few discrete types, our algorithm recovers bounded regret, as in [69] (cf. Corollary 5). For the special class of non-atomic distributions with density bounded away from zero, CWG is identical to CE since there are no gaps and we recover the logarithmic regret scaling result of [73] and [70] (cf. Corollary 4).

(iii) Unifying Algorithm: Returning to general resource allocation problems, we propose a versatile algorithm called **Repeatedly Act using Multiple Simulations (RAMS)**, which offers a practical and data-driven approach to resource allocation. At each  $t$ , RAMS simulates multiple future demand scenarios. Each possible allocation decision at  $t$  results in different cumulative rewards in hindsight, in each demand scenario. RAMS greedily selects the allocation decision which maximizes the average over scenarios of the cumulative reward in hindsight. Unlike previous algorithms, RAMS does not require to be tuned to specific distribution features, and by its design can organically leverage the data-driven simulations of the future which are typically available in practical applications. In terms of performance, we establish a meta result (Theorem 6) that shows that RAMS is guaranteed to inherit the regret performance guarantee of any algorithm satisfying certain conditions (specified in Theorem 6). This result, in conjunction with Theorem 5, implies that RAMS is near-optimal for the multisecretary problem and naturally incorporates the *conservativeness with respect to gaps* principle. Furthermore, our meta theorem, together with existing results on other algorithms in the literature, tells us that RAMS is near-optimal in a variety of settings for NRM and Order Fulfillment problems.

### 2.1.1 Related Literature

The classical secretary problem was introduced by [74] and [75]. The multisecretary variant of the above problem was initially studied by [76] and [77]. Recently, [69] showed that, when the distribution of types is discrete, regret is bounded uniformly for all values of the number of candidates  $T$  and the hiring budget  $B$ , where the constant may scale with the reciprocal of the

minimum probability mass on any type. In order to prove this result, they devise an adaptive policy called the Budget-Ratio (BR) policy where they compare the ratio of the remaining budget to the remaining number of candidates to interview and make the hire/reject decision by comparing the budget ratio to some fixed thresholds. This regret guarantee, in conjunction with a lower bound on regret from [77] yields a tight understanding of the class of distributions supported on a few discrete types. Note that the classical secretary problem and its generalization considered in [77] do not assume the knowledge of the reward distribution. However following the work of [69], the variant of multi-secretary with distributional knowledge has also been referred to as the multi-secretary problem and we will also employ this terminology.

At the other extreme, for a continuum of types, [73, 70] show that instead of the regret being uniformly bounded, the best possible scaling for a certain class of non-atomic distributions with contiguous support is  $\Theta(\log T)$  ([70] shows that this is true for the more general network revenue management problem as well). In the context of the multisecretary problem, they devise a simple threshold policy based on the budget ratio to achieve this regret scaling. However, the class of non-atomic distributions considered in these papers requires the probability density function to be bounded away from zero. In a parallel line of inquiry, the set of distributions examined by [78] bears close resemblance to our own. Yet, there are marked differences in the settings and results. Specifically, [78] concentrate on the auction setting involving a single item and restrict their study to continuous distributions.

The multi-secretary problem is a special case of a broader class of network revenue management (NRM) problems, or more broadly dynamic resource constrained reward collection problems; see [34] for a recent survey and unified modeling framework for this class of problems. There is a wide variety of applications in auction theory [77], online resource allocation [76, 71], order fulfillment [72], among others. Note that this literature typically assumes a small number of types.

[36, 79] generalized the arguments in [69] to a broader class of online packing and online matching problems and proved a uniform regret guarantee across all values of capacity  $B$  and time horizon  $T$ . They developed a technique called *compensated coupling* and used it to prove a constant

regret guarantee without requiring any non-degeneracy assumptions. [37] also proved constant regret guarantees for a class of NRM problems, however their algorithm and proof techniques differ from those of [36, 79]. While all these papers impressively establish constant regret bounds, all of them assume a few discrete types, and their regret bounds scale polynomially in the number of types. However in many practical systems, the number of types is, in fact, large.

Simulation-based algorithms have been studied in the network revenue management literature [38, 39], albeit without any regret guarantees. The idea in these papers is to solve multiple stochastic optimization problems with different realizations instead of a single fluid relaxation and average the shadow prices of the different optimization problems and implement a bid-price control. Recently, [40] and [80] have used related ideas to develop algorithms for online bin packing with a few types.

Another line of research connected to our work is on prophet inequalities, in particular  $k$ -unit prophet inequalities ( $k$  corresponds to the budget  $B$  described earlier). The  $k$ -unit prophet inequality problem, originally studied in [81], analyzes the competitive ratio which is defined as the ratio of the expected performance of an algorithm to the expected performance of the hindsight optimal in the worst case over the reward distributions, where the focus is on deriving tight guarantees in terms of  $k$ . The seminal work of [82] proved a guarantee of  $1 - 1/\sqrt{k+3}$  on the competitive ratio and since then this result has been improved upon by [83] and [84]. One key distinction between this stream and our work is that we consider i.i.d values from a known distribution, which allows to prove stronger guarantees on the regret. The competitive ratio results above would imply a regret scaling of  $\Theta(\sqrt{T})$ , whereas we show that if the distribution is known and i.i.d, it is possible to do better even under the worst-case when the budget  $B$  scales linearly in  $T$  (cf. Theorem 5).

## 2.2 Model

We consider a dynamic resource allocation problem with a *known* finite time horizon  $T$ . There are  $d$  resources and the decision maker is endowed with an initial budget vector  $B \in \mathbb{R}^d$  for the resources. At each time  $t = 1, 2, \dots, T$ , a request  $\theta_t$  is drawn independently from a *type* set  $\Theta$

via some distribution  $F$  which is *known* to the decision maker. Upon observing a request  $\theta_t$ , the decision maker takes an action  $a_t \in \mathcal{A}(B_t, \theta_t)$  where  $\mathcal{A}(B_t, \theta_t)$  is the set of feasible actions at time  $t$  which depends on the remaining budget  $B_t$  and the request  $\theta_t$ . Let  $\mathcal{A} \triangleq \cup_{B \geq 0} \cup_{\theta \in \Theta} \mathcal{A}(B, \theta)$  denote the set of all possible actions. Upon taking an action  $a_t$ , the decision maker collects a reward  $r_t$  which depends on the request  $\theta_t$  and the action  $a_t$ . We denote by  $r : \Theta \times \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$  the reward function. Taking an action consumes resources and the amount of resource consumed depends on the request  $\theta$ , and is denoted by a consumption function  $c : \Theta \times \mathcal{A} \rightarrow \mathbb{R}^d$  where  $c_k(\theta, a)$  is the amount of  $k$ -th resource consumed when the request is  $\theta$  and action is  $a$ . Given a request  $\theta_t$  and action  $a_t$ , the remaining budget is updated as per  $B_{t+1} = B_t - c(\theta_t, a_t)$ ; the action  $a_t$  is required to be such that each coordinate of  $B_{t+1}$  is non-negative. We assume that there is a null action  $a_0 \in \mathcal{A}$  which consumes no resources and generates no reward, i.e.,  $r(\theta, a_0) = 0$  for all  $\theta \in \Theta$  and  $c(\theta, a_0) = 0_{d \times 1}$  for all  $\theta \in \Theta$ . Further, we will assume that  $|r(\theta, a)| \leq 1$  and  $\|c(\theta, a)\|_\infty \leq 1$  for all  $\theta \in \Theta$  and  $a \in \mathcal{A}$ .

A policy is said to be an *online* (non-anticipating) policy if the decision on the  $t$ -th request is based only on the request  $\theta_t$  at time  $t$ , the past requests,  $\{\theta_j\}_{j=1}^{t-1}$  and the history of the actions  $\{a_j\}_{j=1}^{t-1}$  up to the time  $t$ . Let  $U_1, U_2, \dots, U_T$  be a sequence of random variables that are independent and uniformly distributed over  $[0, 1]$  and independent of the requests  $\theta_1, \theta_2, \dots, \theta_T$ . (The  $U$ s will allow us to accommodate randomized policies.) Define the filtration  $\mathcal{F}_t = \sigma(\theta_1, U_1, \theta_2, U_2, \dots, \theta_t, U_t)$  for all  $t \in [T]$ . A feasible online policy  $\pi$  is a sequence of  $\{\mathcal{F}_t : t \in [T]\}$ -measurable random variables  $\{a_1^\pi, a_2^\pi, \dots, a_T^\pi\}$  such that  $\sum_{t=1}^T c(\theta_t, a_t^\pi) \leq B$  almost surely. We define the set of feasible online policies as  $\Pi(B, T)$ . For any feasible and online policy  $\pi \in \Pi(B, T)$ , define  $R_t^\pi = \sum_{k=1}^t r(\theta_k, a_k^\pi), \forall t \in [T]$  to be the accumulated reward up to time  $t$ . The total expected reward under a policy  $\pi \in \Pi(B, T)$  is given by  $V_1^\pi(B, T) = \mathbb{E}[R_T^\pi] = \mathbb{E}[\sum_{t=1}^T r(\theta_t, a_t^\pi)]$ . Fix  $T \in \mathbb{N}$  and  $B \in \mathbb{R}_{\geq 0}^d$ , the objective is to maximize the total expected reward given by  $V_1^*(B, T) = \sup_{\pi \in \Pi(B, T)} V_1^\pi(B, T)$ .

Next we consider the hindsight (hs), full-information version of the problem in which the requests  $\theta_{\geq 1} = \{\theta_1, \theta_2, \dots, \theta_T\}$  are known apriori. In the hindsight setting, the problem essentially

reduces to solving  $V_1^{\text{hs}}(B, T; \theta_{\geq 1}) = \max_a \{\sum_{t=1}^T r(\theta_t, a_t) : (a_1, \dots, a_T) \in |\mathcal{A}|^T \text{ and } \sum_{t=1}^T c(\theta_t, a_t) \leq B\}$  and the total expected value by the hindsight optimal problem is given as  $V_1^{\text{hs}}(B, T) = \mathbb{E} [V_1^{\text{hs}}(B, T; \theta_{\geq 1})]$ . It trivially follows that  $V_1^{\text{hs}}(B, T) \geq V_1^\pi(B, T), \forall \pi \in \Pi(B, T)$  and  $\forall B \in [T]$ . To measure the performance of a feasible online policy  $\pi \in \Pi(B, T)$ , we consider the hindsight problem as a benchmark and define the (expected) regret of the policy  $\pi$  as the difference between the expected value of the hindsight problem and the expected value attained by the policy  $\pi$  i.e.,  $\text{Regret}(B, T; \pi) \triangleq V_1^{\text{hs}}(B, T) - V_1^\pi(B, T)$ . We also define the (minimum achievable, expected) regret as the difference between the expected value of the hindsight problem and the expected value under the optimal online policy  $\pi^* \in \Pi(B, T)$ .

$$\text{Regret}(B, T) = \inf_{\pi \in \Pi(B, T)} \text{Regret}(B, T; \pi) = V_1^{\text{hs}}(B, T) - V_1^*(B, T).$$

In what follows, we will focus on characterizing the growth rate of  $\text{Regret}(B, T)$  as a function of  $T$  and the characteristics of the distribution of types.

Next we discuss the three important classes of online resource allocation problems.

**Network Revenue Management.** In this problem each request  $\theta = (r_\theta, \mathbf{c}_\theta)$  is presented with a single reward  $r_\theta \geq 0$  and a consumption vector  $\mathbf{c}_\theta \in \mathbb{R}^d$ . We have that  $\mathcal{A} = \{a_0 = \text{reject}, a_1 = \text{accept}\}$ . The reward and consumption functions are given as

$$\begin{aligned} r(\theta, \text{reject}) &= 0, & c(\theta, \text{reject}) &= \mathbf{0}_{d \times 1} \\ r(\theta, \text{accept}) &= r_\theta, & c(\theta, \text{accept}) &= \mathbf{c}_\theta. \end{aligned}$$

**Online Matching (Order Fulfillment).** In this problem each request  $\theta = r_\theta$  is presented with a vector of rewards  $\mathbf{r}_\theta \in \mathbb{R}^d$ . Each request wants to consume at most one unit of any single resource. The action set is  $\mathcal{A} = \{a_0, a_1, \dots, a_d\}$  where  $a_k$  denotes that the request is matched to resource  $k$  with  $a_0$  being the null action denoting that the request is rejected. The reward and consumption

functions are given as

$$r(\theta, a_0) = 0, \quad c(\theta, a_0) = \mathbf{0}_{d \times 1}$$

$$r(\theta, a_k) = \mathbf{r}_{\theta, k}, \quad c(\theta, a_k) = \mathbf{e}_k, \quad \forall k \in \{1, 2, \dots, d\}$$

where  $\mathbf{r}_{\theta, k}$  denotes the  $k$ -th coordinate of  $\mathbf{r}_\theta$  and  $\mathbf{e}_k$  is a  $d$ -dimensional vector with the  $k$ -th coordinate being one and all other coordinates being zero.

**Multi-secretary Problem.** For the case of one resource ( $d = 1$ ), network revenue management and online matching are equivalent problems and this special case is referred to as the multi-secretary problem. We have that  $c(\theta, \text{accept}) = 1$  for all  $\theta \in \Theta$ . In the context of the multi-secretary problem, the request type (equivalently, reward) will be referred to as the candidate ability.

### 2.3 Fundamental Limits on Achievable Performance

To delve deeper into the intrinsic drivers of performance, we initially focus on the multi-secretary problem – a cornerstone model in online resource allocation. Clearly, any lower bound established for the multi-secretary problem directly translates into a lower bound for a broader range of online resource allocation problems like NRM and online matching. We now define two classes of distributions under which the multi-secretary problem has been previously studied.

**Assumption 2 (Small Number of Types)** *The type (reward) distribution  $F$  is supported on a finite set and the rewards are assumed to be in the interval  $[0, 1]$ .*

**Remark 5** *Many prior works refer to this as the “finite types setting”, and establish constant regret guarantees [see, e.g., 69, 37, 36]. However, these guarantees scale linearly with the number of types. Hence, they are most relevant when the size of discrete types set is small. To emphasize this aspect, we use the phrases “small number of types” or “small discrete set” or “few types” to describe this setting.*

**Assumption 3 (Infinitely Many Types with density bounded away from zero)** *The type (reward) distribution  $F$  is supported on an infinite set and  $F$  admits a density  $f$  which is bounded from below and above, i.e., there exist  $0 < \underline{v} \leq \bar{v} < \infty$  such that  $\underline{v} \leq f(\theta) \leq \bar{v}$  for all  $\theta \in \Theta$ . The rewards are assumed to be in the interval  $[0, 1]$ .*

To interpolate between these two class of distributions, we will introduce a general class of distributions which will capture the distributions with a few types and infinitely many types with bounded density as special cases.

### 2.3.1 General Class of Distributions For the multisecretary Problem

We will anchor our analysis around a general family of distributions which allow for gaps in the type space and can capture as special cases discrete distributions as well as the non-atomic distributions with density uniformly bounded away from zero. We call this family  $(\beta, \varepsilon_0, \delta)$ -clustered distributions. For any  $q \in [0, 1]$ , we define  $F^{-1}(q) \triangleq \inf\{v : F(v) \geq q\}$ .

**Definition 2 ( $(\beta, \varepsilon_0, \delta)$ -clustered distributions)** *Fix  $\beta \in [0, \infty)$ ,  $\varepsilon_0 \in (0, 1]$  and  $\delta \in [0, 1]$ . A distribution  $F$  is said to be  $(\beta, \varepsilon_0, \delta)$ -clustered if there exists  $n \in \mathbb{N} \cup \{0\}$  and gap quantiles  $q_0^* = 0 < q_1^* < \dots < q_n^* < q_{n+1}^* = 1$  such that we have*

(a) *(Generalized cluster “density” requirement)  $\forall i \in [n + 1], \forall q, \tilde{q} \in (q_{i-1}^*, q_i^*]$ , we have that*

$$|F^{-1}(q) - F^{-1}(\tilde{q})| \leq C|q - \tilde{q}|^{\frac{1}{\beta+1}} + \delta \text{ for some constant } C < \infty.$$

(b) *(Cluster size requirement)  $q_i^* - q_{i-1}^* \geq \varepsilon_0, \forall i \in [n + 1]$ .*

Let  $\mathcal{F}_{\beta, \varepsilon_0, \delta}$  denote the class of  $(\beta, \varepsilon_0, \delta)$ -clustered distributions. This class includes a wide variety of distributions. An important sub-class is the one with  $\delta = 0$ , which we denote by  $\mathcal{F}_{\beta, \varepsilon_0}$ . We refer to distributions in this subclass as  $(\beta, \varepsilon_0)$ -clustered.

Define  $H_i \triangleq [F^{-1}((q_{i-1}^*)^+), F^{-1}(q_i^*)]$ , for all  $i \in [n + 1]$ , where  $F^{-1}(q^+) \triangleq \lim_{\epsilon \rightarrow 0^+} F^{-1}(q + \epsilon)$ . We will refer to the  $(H_i)$ 's as *mass clusters* or just *clusters*. We will use the term *gaps* to refer to the complementary intervals  $G_i \triangleq (F^{-1}(q_i^*), F^{-1}((q_i^*)^+))$  for  $i \in [n]$ , and the intervals at

the extremes  $G_0 = [0, F^{-1}(0^+))$ ,  $G_{n+1} = (F^{-1}(1), 1]$ , since they contain no probability mass. The requirement (a) can be thought of as a within-cluster “density” requirement, which becomes weaker as  $\beta$  increases; we can think of  $\beta$  as quantifying the within-cluster mass density (with a decreasing relationship). When  $\delta = 0$ , this requirement corresponds to  $F^{-1}$  being  $(1/(\beta + 1))$ -Hölder continuous on the mass clusters. Requirement (b) is a cluster size requirement,  $\varepsilon_0$  being the minimum cluster size; this requirement becomes more stringent as  $\varepsilon_0$  increases. The parameter  $\delta$  provides us with additional flexibility in modelling our distributions. One such practically relevant class of distributions is the one with a large number of discrete types, which can be modelled using the parameter  $\delta$  (cf. Example 4). We believe that a very broad set of general distributions can be either represented or approximated using a  $(\beta, \varepsilon_0, \delta)$ -clustered distribution (cf. Appendix B.8 for an example of singular distribution). In general, there is some flexibility on how the distributions are modelled, more specifically how the types are aggregated into clusters, and this is associated with a tradeoff between  $\delta$  and  $\varepsilon_0$  (and potentially  $\beta$ ). Please refer to Appendix B.9 for more details.

Next we present some examples of  $(\beta, \varepsilon_0, \delta)$ -clustered distributions including discrete distributions, as well the uniform distribution, along with the appropriate choices of gap quantiles.

**Example 1 (Discrete Distributions)** *Consider a discrete distribution [as studied in 69]. Let the support be  $\{\theta_1, \theta_2, \dots, \theta_n\}$  with probability masses  $\{p_1, p_2, \dots, p_n\}$ . Assume that  $0 \leq \theta_1 < \theta_2 < \dots < \theta_n \leq 1$ . We make use of the natural choice of gap quantiles  $q_i^* = \sum_{j=1}^i p_j$  for all  $i \in [n - 1]$ , leading to gaps  $G_0 = [0, \theta_1)$ ,  $G_i = (\theta_i, \theta_{i+1}) \forall i \in [m - 1]$ ,  $G_n = (\theta_n, 1]$  and clusters  $H_i = \{\theta_i\} \forall i \in [n]$ . Now for  $q, \tilde{q} \in (q_{i-1}^*, q_i^*) = Q_i$ , we have that  $|F^{-1}(q) - F^{-1}(\tilde{q})| = 0 \leq |q - \tilde{q}|$ , i.e., the cluster density requirement is satisfied for  $\beta = 0$  and  $\delta = 0$ . Defining  $\varepsilon_0 \triangleq \min\{p_1, p_2, \dots, p_n\}$  the cluster size requirement is satisfied. Therefore the discrete distribution belongs to the class of  $(0, \varepsilon_0)$ -clustered distributions where  $\varepsilon_0$  is the minimum probability mass in the support.*

**Example 2 (Non-atomic Distributions with Contiguous Support)** *Consider the non-atomic distributions with pdf  $f$  considered in [70] (Assumption 3). Assume that there exists  $\alpha_0 > 0$  such that  $f(x) \geq \alpha_0, \forall x \in [0, 1]$ . (The uniform distribution over  $[0, 1]$  is a special case of these distributions with  $f(x) = 1$  for all  $x \in [0, 1]$ .) Such distributions are  $(\beta = 0, \varepsilon_0 = 1, \delta = 0)$ -clustered*

distributions with  $n = 0$  gaps, i.e.,  $F^{-1}$  is 1-Hölder continuous over the interval  $(0, 1]$  with the constant  $C = 1/\alpha_0$ . The gap quantiles are only the trivial ones  $q_0^* = 0$  and  $q_1^* = 1$ . There is a single mass cluster  $H_1 = [0, 1]$  with mass 1, which clearly satisfies the cluster density requirement with  $\beta = 0, \varepsilon_0 = 1$  and  $\delta = 0$ .

**Example 3 (A class of bimodal distributions)** An example of a  $(\beta, \varepsilon_0)$ -clustered distribution with  $n = 1$  gap (with gap quantile  $q_1^* = 1/2$ ), for general  $\beta \geq 0$  and  $\varepsilon_0 = 1/2$ , which we will make use of to prove our lower bound results is presented below:

$$F_\beta(x) = \begin{cases} -2 \cdot 4^\beta \cdot \left(\frac{1}{4} - x\right)^{\beta+1} + \frac{1}{2} & 0 \leq x \leq \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} \leq x \leq \frac{3}{4} \\ 2 \cdot 4^\beta \cdot \left(x - \frac{3}{4}\right)^{\beta+1} + \frac{1}{2} & \frac{3}{4} \leq x \leq 1 \end{cases}, \quad F_\beta^{-1}(q) = \begin{cases} \frac{1-(1-2q)^{\frac{1}{\beta+1}}}{4}, & 0 \leq q \leq \frac{1}{2} \\ \left[\frac{1}{4}, \frac{3}{4}\right] & q = \frac{1}{2} \\ \frac{(2q-1)^{\frac{1}{\beta+1}}+3}{4}, & \frac{1}{2} < q \leq 1 \end{cases} \quad (2.1)$$

It is easy to see that  $F_\beta^{-1}$  in (2.1) is a  $(\beta, 1/2)$ -clustered distribution, with one gap  $G_1 = (1/4, 3/4)$  and clusters  $H_1 = [0, 1/4]$  and  $H_2 = [3/4, 1]$ . Refer to Figure 2.1 for a plot of the density  $f_\beta$  and the CDF  $F_\beta$  of the  $(\beta, 1/2)$ -clustered distribution defined in (2.1).

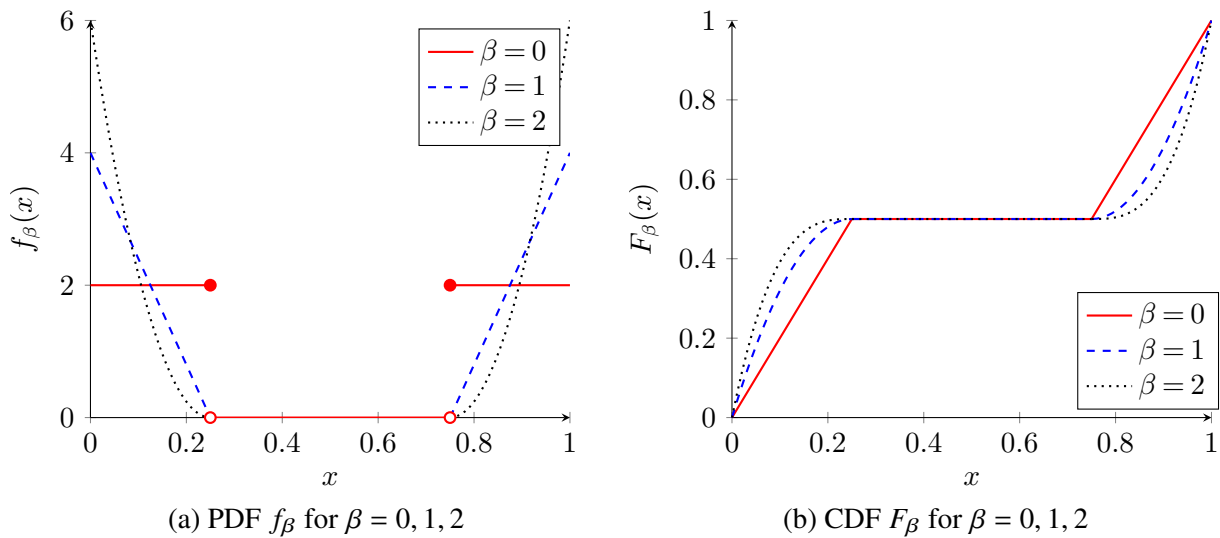


Figure 2.1: The PDF and CDF for the bimodal distributions  $F_\beta$ . Notice the gap from  $1/4$  to  $3/4$ .

Observe that  $(\beta, \varepsilon_0)$ -clustered distributions already allow us to capture not only the previously studied distributions such as distributions with few types and continuous distributions (with density bounded below), but also a mixture of atomic and non-atomic distributions with gaps. As mentioned previously, the parameter  $\delta$  provides us with additional flexibility to model distributions with a large number of discrete types, which may be of practical relevance. One such example is that of *many small discrete types* which we present below.

**Example 4 (Many Small Discrete Types)** *Fix a small  $\delta > 0$  and consider a discrete distribution with many small types supported on the points  $\mathcal{S} = \{0, \delta, 2\delta, \dots, 1/4\} \cup \{3/4, 3/4 + \delta, 3/4 + 2\delta, \dots, 1\}$  with probability mass  $2\delta$  on each of the points in  $\mathcal{S}$ . This constitutes a setting with many small discrete types since there are a large number of atomic types (separated by small empty intervals) and the probability mass of each type is small, i.e., it is proportional to  $\delta$ . This instance of many small discrete types captures the salient feature of the order fulfillment problem that there are a large number of demand types (e.g., zipcodes) with each demand type having small probability mass and these demand types are spatially clustered with possibly large gaps between different clusters of demand types. As  $\delta \rightarrow 0$ , we recover the bimodal uniform distribution  $F_0$  in the limit. One can similarly consider similar many-small-discrete-type analogs for other  $(\beta, \varepsilon_0)$ -clustered distributions. Note that the many small types need not be uniformly spaced. We require that the maximum distance between the discretized grid points be at most  $\delta$ . In such discretizations, we have some flexibility in choosing which empty intervals to classify as “gaps”. In the case of many small types, if the size of the empty intervals (due to discretization) is at most  $\delta$  then we may consider the entire clump of these many small types as belonging to one cluster (say,  $H_i$ ) and hence one quantile interval ( $Q_i$ ).*

### 2.3.2 Fundamental Lower bound on Performance

In this section we present a novel driver of regret scaling: the shape of the candidate ability (or value) distribution around gaps which is characterized by the parameter  $\beta \in [0, \infty)$  and show that for  $\beta > 0$ , polynomial regret scaling is unavoidable. To focus on the scaling with parameter  $\beta$ , we

fix  $\delta = 0$  and  $\varepsilon_0 = 1/2$ .

**Theorem 4 (universal lower bound)** *Fix  $\delta = 0, \varepsilon_0 = 1/2$  and consider any  $\beta \in [0, \infty)$ . Then there exists a candidate-ability distribution  $F \in \mathcal{F}_{\beta, \varepsilon_0}$ , a time horizon  $T_0 < \infty$ , a universal constant  $c > 0$  such that, for all  $T \geq T_0$  and for any online policy  $\pi \in \Pi(B, T)$ , we have that*

$$\sup_{B \in [T]} \text{Regret}(B, T; \pi) \geq (c/(1 + \beta))T^{\frac{1}{2} - \frac{1}{2(\beta+1)}} \mathbb{1}\{\beta > 0\} + c \log T \mathbb{1}\{\beta = 0\}.$$

This theorem provides an impossibility result: it says that for any fixed  $\beta \in [0, \infty)$ , there exists a distribution for which no online policy can achieve a better regret scaling than the one presented in Theorem 4. This lower bound also highlights that the fundamental limits of the regret scaling are governed by the parameter  $\beta$  which characterizes the curvature of the distribution around the gap boundaries. We observe that as  $\beta \rightarrow \infty$ , the scaling of regret approaches  $\sqrt{T}$ ; i.e., no matter the online policy, it will suffer regret nearly as large as that of a simple non adaptive policy. Hence  $\beta$  can be seen as characterizing the “hardness” of an instance. The parameter  $\beta$  has a physical interpretation as well. It captures how mass accumulates in the type space. For some intuition, consider the  $F_\beta$  distribution described in (2.1) and consider the gap boundary at  $3/4$ . As we move from the boundary point  $3/4$  to a distance  $\delta$  into the adjacent cluster, i.e., to  $3/4 + \delta$ , the probability mass accrued grows as  $C\delta^{\beta+1}$  for some universal constant  $C > 0$ . Alternately, to accrue a probability mass of  $\varepsilon$ , we need to move a distance  $C\varepsilon^{\frac{1}{\beta+1}}$  from the boundary  $3/4$  into the adjacent cluster. Therefore as  $\beta$  increases, the distance one needs to travel to collect a probability mass of  $\varepsilon$  also increases and this property is what makes the instances harder as  $\beta$  increases.

For  $\beta = 0$ , our lower bound follows from [70]. To establish our bound for  $\beta > 0$ , we consider the distributions  $F_\beta$  defined in (2.1). At a high level, we consider two events of  $\Omega(1)$  probability – one is a perturbation of the other – under one event (denoted as  $\mathcal{H}$ ), there are more than the expected number of arrivals with values at least  $3/4$  (“high” types) and hence the hindsight threshold is (slightly) more than  $3/4$ , and on the other event (denoted as  $\mathcal{L}$ ), there are fewer than expected number of arrivals with value at least  $3/4$  and hence the hindsight threshold is (slightly) less than

1/4. While the hindsight optimal policy does well on both the events, the optimal online policy can only do well on one or the other but not in both. We show that any online algorithm must make at least  $\Omega(\sqrt{T})$  mistakes on at least one of the two events, and leveraging how the mass accumulates over space (characterized by Definition 2), one may show that the cost of each of these mistakes is  $\Omega(T^{-\frac{1}{2(\beta+1)}})$ . Combining the two gives us that the cumulative regret scales as  $\Omega(T^{\frac{1}{2}-\frac{1}{2(\beta+1)}})$ . We elaborate on this in the formal proof in Appendix B.1.

## 2.4 Algorithmic Design Principles for Near Optimal Performance

Having established a spectrum of fundamental performance boundaries, it is natural to inquire if it is possible to achieve these limits, and if so, what algorithms are capable of attaining these fundamental limits. A prevalent algorithmic principle in the network revenue management literature is the Certainty Equivalent (CE) heuristic. This approach solves a deterministic approximation of a stochastic optimization problem by substituting random variables with their expected values. Given its widespread use, the CE heuristic emerges as a natural initial candidate for analysis and characterization of achievable performance. In this section, we will focus on the CE heuristic for non-atomic distributions to avoid any tie-breaking issues which are present for atomic distributions. For the multisecretary problem, the CE heuristic is defined as follows: at each time  $t$  (before the arrival of request  $\theta_t$ ), given a remaining budget  $B_t$  and remaining number of time steps  $T - t + 1$ , we compute the budget ratio  $B_t/(T - t + 1)$  and accept the request  $\theta_t$  if and only if  $r(\theta_t, \text{accept}) \geq F^{-1}(1 - B_t/(T - t + 1))$ . Note that the CE heuristic employs an adaptive threshold at each time  $t$ .

### 2.4.1 Failure of the CE policy under many types with gaps

Indeed, in the case of non-atomic distributions with density uniformly bounded away from zero, [73] and [70] showed that CE achieves  $O(\log T)$  regret, and that this is the best scaling achievable. However, it turns out that as soon as one introduces a gap in these non-atomic distributions (as in Example 3), the performance of CE degrades significantly. This phenomenon is documented in

the proposition below.

**Proposition 3 (Failure of CE)** *Fix any  $\eta \in (0, 1)$  and  $\varepsilon \in (0, 1/2]$ . Suppose the candidate-ability distribution  $F$  is any non-atomic distribution that has a gap of length at least  $\eta$ , i.e.,  $\exists c \in (0, 1 - \eta)$  such that  $F(c) = F((c + \eta)^-)$ , and such that there is mass at least  $\varepsilon$  on each side of the gap, i.e.,  $\min\{F(c), 1 - F(c)\} \geq \varepsilon$ . Then for the CE policy, there exists  $T_0 \equiv T_0(\varepsilon) < \infty$ , a constant  $c \equiv c(\eta, \varepsilon) > 0$  and  $B \in [T]$  such that  $\text{Regret}(B, T; \text{CE}) \geq c\sqrt{T}$  for all  $T \geq T_0$ .*

The regret of the CE policy increases dramatically if there is a gap in the types, even when one maintains the uniform distribution of types (or any other distribution) outside of the gap. As a matter of fact, the regret scaling is as large as that of a non-adaptive policy. The result in Proposition 3 is analogous to the results for few types in the literature. The main driver of  $\Omega(\sqrt{T})$  regret scaling for both the many types with gaps and finite types settings is *degeneracy*, i.e., situations where the dual variables corresponding to the initial fluid model LP are not unique. This issue is well documented in the setting with finitely many types [37, 36], but also manifests in the case of non-atomic distributions with gaps. As such the proof of Proposition 3 follows from the proof of the analogous result for finitely many types in [37, Proposition 2].

#### 2.4.2 Conservativeness with respect to gaps

We observed that the CE policy breaks down for distributions with many types and “gaps” (intervals) of absent types; it suffers  $\Omega(\sqrt{T})$  regret, as large as that of a non-adaptive algorithm. We identified that the main driver for the  $\Omega(\sqrt{T})$  regret of the CE policy is the presence of gaps. To solve this issue, we introduce a new algorithmic principle which we call “conservativeness with respect to gaps” (CWG), and use it to provably achieve near optimal regret scalings for the  $(\beta, \varepsilon_0, \delta)$ -clustered distributions which allow for gaps. The idea of CWG is that if there is a risk that the acceptance threshold based on CE will move across a given gap in the future, then CWG uses that gap as the acceptance threshold instead of using the CE-based threshold. Based on the CWG principle, we devise a new policy with the same name, which we present in Algorithm 2.

---

**Algorithm 2:** Conservativeness with respect to Gaps (CwG)

---

**Input:** Time Horizon  $T$ , Hiring Budget  $B$ ,  $(\beta, \varepsilon_0, \delta)$ -clustered dist.  $F$  with gaps

$$G_i = (a_i, b_i).$$

**Initialize:**  $B_1 = B, q_i^* = F(a_i) = F(b_i^-), \forall i \in [n], \tilde{T} = \max\{0, T - \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor\}$

**for**  $t = 1$  **to**  $\tilde{T}$  **do**

$$p_t^{\text{CE}} = 1 - \frac{B_t}{T-t+1}$$

$$\mathcal{S}_t = \left\{ i : p_t^{\text{CE}} \in \mathcal{B} \left( q_i^*, \sqrt{\frac{2 \log(T-t+1)}{T-t+1}} \right) \right\}$$

**if**  $\mathcal{S}_t = \emptyset$  **then**

$$p_t^{\text{CwG}} = p_t^{\text{CE}}$$

**else**

$$j_t^* = \arg \min_{i \in \mathcal{S}_t} |p_t^{\text{CE}} - q_i^*|$$

$$p_t^{\text{CwG}} = q_{j_t^*}^*$$

**end**

Observe a candidate of ability  $\theta_t$  and form the set  $\mathcal{I}_t = \{q \in [0, 1] : F^{-1}(q) = \theta_t\}$

Let  $X_t$  be a uniform sample from the set  $\mathcal{I}_t$

**if**  $X_t \geq p_t^{\text{CwG}}$  **and**  $B_t > 0$  **then**

Hire the candidate and  $B_{t+1} \leftarrow B_t - 1$

**else**

Reject the candidate and  $B_{t+1} \leftarrow B_t$

**end**

**end**

Define  $p_{\tilde{T}+1}^{\text{CE}} = 1 - \frac{B_{\tilde{T}+1}}{T-\tilde{T}}$

**for**  $t = \tilde{T} + 1$  **to**  $T$  **do**

Observe a candidate of ability  $\theta_t$  and form the set  $\mathcal{I}_t = \{q \in [0, 1] : F^{-1}(q) = \theta_t\}$

Let  $X_t$  be a uniformly random sample from the set  $\mathcal{I}_t$

**if**  $X_t \geq p_{\tilde{T}+1}^{\text{CE}}$  **and**  $B_t > 0$  **then**

Hire the candidate and  $B_{t+1} \leftarrow B_t - 1$

**else**

Reject the candidate and  $B_{t+1} \leftarrow B_t$

**end**

**end**

---

The algorithm operates in two phases. For simplicity, assume that  $T \geq \lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$ . We begin by describing the first phase. For the first  $\tilde{T} \triangleq T - \lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$  steps the algorithm uses the CWG principle, where if the re-solving threshold  $p_t^{\text{CE}}$  is close to a gap, we modify it by instead using the quantile corresponding to the boundary of the gap as our acceptance threshold  $p_t^{\text{CWG}}$ . It remains to clarify how the quantile threshold  $p_t^{\text{CWG}}$  translates to an accept/reject decision for the arrival at  $t$ . After observing the type  $\theta_t$ , we form the set of corresponding quantiles  $\mathcal{I}_t$ . If  $\mathcal{I}_t$  is a singleton (this is the case if  $\theta_t$  does not lie at an atom of  $F$ ) then we have that its unique element  $X_t = F(\theta_t)$ . If  $\theta_t$  lies at an atom of  $F$ , the set  $\mathcal{I}_t$  is a corresponding interval (recall Example 1). If  $p_t^{\text{CWG}} \notin \mathcal{I}_t$  then the hire/reject decision is unambiguous. The only case of ambiguity is  $p_t^{\text{CWG}} \in \mathcal{I}_t$ . To handle this case, we make use of randomization to break ties by drawing  $X_t$  uniformly from the interval  $\mathcal{I}_t$ , and hiring the candidate only if the  $X_t$  is weakly greater than  $p_t^{\text{CWG}}$ .

We now describe the second phase of the algorithm. In the final  $\lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$  time steps, the radius  $\sqrt{2 \log \tau / \tau}$  (where  $\tau = T - t + 1$  is the number of remaining time steps) by which we measure the closeness of CE threshold  $p_t^{\text{CE}}$  and the gap quantiles  $\{q_i^*\}_{i=1}^n$  becomes too large, i.e.  $\sqrt{2 \log \tau / \tau} > \varepsilon_0/2$ . This results in more than one gap quantiles being in the  $\sqrt{2 \log \tau / \tau}$  neighborhood of  $p_t^{\text{CE}}$  which in turn makes the choice of  $p_t^{\text{CWG}}$  ambiguous and further complicates the regret analysis. In order to avoid this ambiguity and simplify the analysis, we employ a static allocation policy in the second phase: we solve for the certainty equivalent threshold  $p_{\tilde{T}+1}^{\text{CE}}$  at time  $\tilde{T} + 1$ , and use that threshold for the remaining  $\lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$  time steps.

## Performance Analysis

**Theorem 5** *For any  $\beta \in [0, \infty)$ ,  $\varepsilon_0 \in (0, 1]$  and  $\delta \in (0, 1]$ , suppose the candidate-ability distribution  $F$  with associated gaps is  $(\beta, \varepsilon_0, \delta)$ -clustered. Then for all  $T \in \mathbb{N}$  and for all  $B \in [T]$ , there*

exists a universal constant  $C < \infty$  such that the regret of the **CWG** policy is upper bounded as

$$\begin{aligned}
\text{Regret}(B, T; \text{CWG}) &\leq \underbrace{C(1 + 1/\beta)(\log T)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} T^{\frac{1}{2} - \frac{1}{2(\beta+1)}} \cdot \mathbb{1}\{\beta > 0\} + C(\log T)^2 \mathbb{1}\{\beta = 0\}}_{(\spadesuit)} \\
&\quad + \underbrace{C\delta\sqrt{T \log T}}_{(\diamondsuit)} + \underbrace{C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0}_{(\heartsuit)}. \tag{2.2}
\end{aligned}$$

*Discussion of Theorem 5.* The regret upper bound can be decomposed as shown in (2.2), where each of the terms has a different driver. The terms in  $(\spadesuit)$  are driven by the shape of the reward distribution around gaps and is characterized by the parameter  $\beta \in [0, \infty)$ . Comparing the term  $(\spadesuit)$  to the lower bound in Theorem 4, we note that the scaling of the upper bound matches the scaling of the lower bound in  $T$  up to a polylogarithmic factor and hence the proposed **CWG** policy is near-optimal. In the case of the **CE** policy, we had identified that the main driver of its worst case regret of  $\Theta(\sqrt{T})$  was the presence of gaps in the distribution of candidate abilities. Theorem 5 tells us that one can overcome the difficulty introduced by gaps in the distribution by using the **CWG** principle that we devised. The term in  $(\diamondsuit)$  is driven by the parameter  $\delta$  which allows us to model distributions with many *small* discrete types (cf. Example 4). We will typically assume that  $\delta$  is small and may scale as  $o(1/\sqrt{T})$ . Note that for the extreme cases of a few types (cf. Example 1) or continuous distributions (cf. Example 2), we have that  $\delta = 0$  and hence the term in  $(\diamondsuit)$  disappears. The term in  $(\heartsuit)$  is driven by the minimum probability mass  $\varepsilon_0$  and is typically assumed to be a constant in  $(0, 1]$ . The contribution of  $(\heartsuit)$  is attributable to the regret accrued due to the static allocation rule employed in Algorithm 2 in the last  $\lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$ . In terms of scaling of  $(\heartsuit)$ , it matches up to polylogarithmic factors the lower bound on regret scaling of  $\Omega(1/\varepsilon_0)$  presented in Lemma 1 of [69].

**Corollary 4** *Suppose the candidate-ability distribution is  $F_0$  where  $F_0$  is as defined in (2.1) with  $\beta = 0$ . Then we have that for all  $T \in \mathbb{N}$  and for all  $B \in [T]$  the regret of our **CWG** policy is upper bounded as  $\text{Regret}(B, T; \text{CWG}) \leq C(\log T)^2$  for the universal constant  $C < \infty$  in Theorem 5.*

**Discussion of Corollary 4.** This corollary follows immediately from Theorem 5 by setting  $\beta = 0$  and  $\delta = 0$ . The distribution  $F_0 = \text{Uniform}([0, 1/4] \cup [3/4, 1])$  is a natural variant of the uniform distribution with a gap. Corollary 4 shows that regret of CWG scales as  $O((\log T)^2)$  for the distribution  $F_0$ . This is a significant improvement on the  $\Omega(\sqrt{T})$  regret scaling of the CE policy for the same distribution  $F_0$ , and the regret of the CWG policy for the  $F_0$  distribution is only a  $\log T$  factor larger than the regret for the uniform distribution. The key takeaway from Corollary 4 in conjunction with Proposition 3 is that the presence of gaps is not a fundamental driver of the achievable regret performance, and one can overcome the difficulty posed by gaps by using the CWG principle.

**Corollary 5 (Constant Regret for discrete distributions)** *Suppose the candidate-ability distribution is  $F$  where  $F$  is a discrete distribution as described in Example 1. Then, for all  $T \in \mathbb{N}$  and for all  $B \in [T]$ , we have  $\text{Regret}(B, T; \text{CWG}) \leq C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0$  for a universal constant  $C < \infty$ .*

**Remark 6** *The discrete distribution considered in Example 1 belongs the class of  $(0, \varepsilon_0)$ -clustered distributions and hence from Corollary 4, it follows that  $\text{Regret}(B, T) = O((\log T)^2)$ . However, recall from Example 1 that for discrete distributions we have that  $\forall i \in [n + 1], \forall q, \tilde{q} \in Q_i, |F^{-1}(q) - F^{-1}(\tilde{q})| = 0$ , and this distinguishes discrete distributions from general  $(0, \varepsilon_0)$ -clustered distributions. This distinction allows us to obtain stronger regret guarantees than the one implied by Corollary 4 and recover the result of [69]. The proof of Corollary 5 follows by modifying the analysis leading to Theorem 5. The modifications enable us to eliminate the  $C(\log T)^2$  term in the regret bound in Theorem 2. We defer the details to Appendix B.3.*

**Corollary 6 (Regret for non-atomic distribution with contiguous support)** *For any  $\beta \in [0, \infty)$ ,  $\varepsilon_0 = 1$ , and  $\delta = 0$ , suppose the candidate-ability distribution  $F$  is  $(\beta, \varepsilon_0 = 1, \delta = 0)$ -clustered ( $F$  has no non-trivial gaps). Then for all  $T \in \mathbb{N}$  and for all  $B \in [T]$ , there exists a universal constant*

$C < \infty$  such that the regret of our CwG policy is

$$\text{Regret}(B, T; \text{CwG}) \leq C \left(1 + \frac{1}{\beta}\right) T^{\frac{1}{2} - \frac{1}{2(1+\beta)}} \mathbb{1}\{\beta > 0\} + C \log T \mathbb{1}\{\beta = 0\}$$

**Discussion of Corollary 6:** This corollary follows immediately from Theorem 5 by setting  $\varepsilon_0 = 1$  and  $\delta = 0$ , except for some polylogarithmic factors. The class of  $(\beta, \varepsilon_0 = 1, \delta = 0)$ -clustered distributions allows for the pdf  $f$  to be zero at some points. An example of such a distribution is given by  $\tilde{F}_\beta(x) = \left(0.5 - 2^\beta (0.5 - x)^{\beta+1}\right) \mathbb{1}\{x \leq 0.5\} + \left(0.5 + 2^\beta (x - 0.5)^{\beta+1}\right) \mathbb{1}\{x > 0.5\}$  where the pdf  $f$  is zero at  $x = 0.5$ . Since there are no non-trivial gaps for the distribution  $\tilde{F}_\beta$ , we choose to treat the whole interval  $[0, 1]$  as a single cluster and hence have  $\varepsilon_0 = 1$ . It can be easily verified that  $\tilde{F}_\beta$  satisfies the “cluster density requirement” in Definition 2 with  $\delta = 0$ . Note that the distribution  $\tilde{F}_\beta$  is not admissible under the assumptions of [70] for  $B = T/2$  and  $\beta > 0$ . Since there are no gaps of positive length in  $(\beta, 1)$ -clustered distributions, the CwG policy boils down to the CE policy. If the probability density function  $f$  is bounded below by a constant, we have  $\beta = 0$  and we recover the  $O(\log T)$  scaling in [70]. If  $f$  is zero at some points, then the regret scaling is determined by  $\beta$  which quantifies how the mass accumulates around types where  $f$  is zero. This result, in conjunction with Theorem 4, proves that the CE policy is near-optimal in the absence of non-trivial gaps.

### 2.4.3 Achieving Conservativeness with respect to Gaps via a Simulation-based Policy

In Algorithm 2, if the re-solving threshold  $p_t^{\text{CE}}$  at time  $t = T - \tau + 1$  was within  $\sqrt{2 \log \tau / \tau}$  of a gap, we modified it as by instead using the quantile corresponding to the boundary of the gap as our acceptance threshold. An alternative to this method is a simulation-based approach, which we’ll outline next, followed by a full treatment in the next section.

Consider the bimodal uniform distribution, described by (2.1) with  $\beta = 0$ . Assume the CE threshold at time  $t$ , denoted as  $p_t^{\text{CE}}$ , is  $1/2 - \epsilon$ , where  $\epsilon$  is sufficiently small ( $\epsilon < \sqrt{2 \log \tau / \tau}$ , where  $\tau = T - t + 1$ ). Under Algorithm 2, the CwG quantile threshold is set to  $p_t^{\text{CwG}} = 1/2$ . Consequently,

only abilities with values of at least  $3/4$  will be accepted at time  $t$ . This is illustrated in Figure 2.2a, where the threshold shifts from  $F^{-1}(p_t^{\text{CE}}) = 1/4 - 2\epsilon$  (in red) to  $F^{-1}(p_t^{\text{CwG}}) = 1/4$  (in blue).

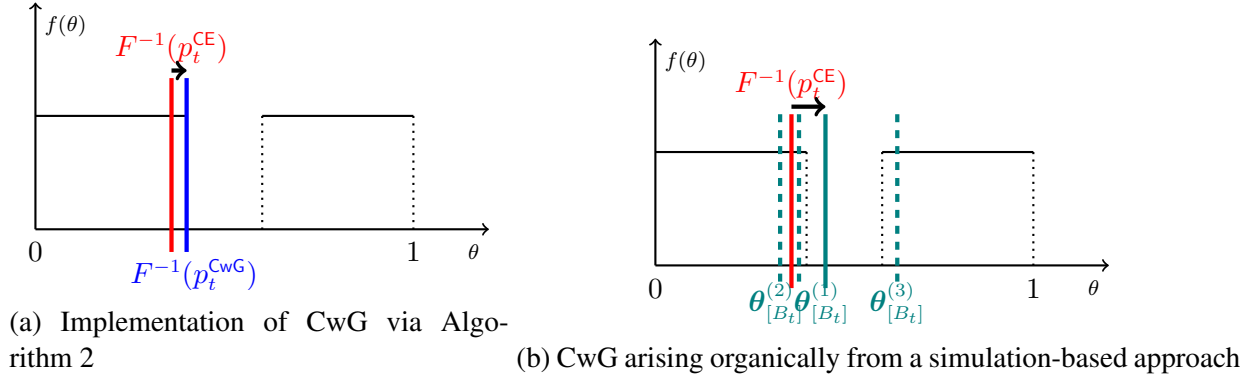


Figure 2.2: Implementation of the CWG principle using two algorithmic approaches.

On the other hand, consider the following simulation-based approach: simulate multiple future demand scenarios. For the  $i$ -th simulated scenario, let  $\theta_{[B_t]}^{(i)}$  denote the value of the  $B_t$ -th largest candidate ability on the simulated sample path  $\tilde{\theta}_{\geq t+1}^{(i)}$ , where  $B_t$  is the remaining budget. The candidate with ability  $\theta_t$  is accepted if  $\theta_t \geq K_t^{-1} \sum_{k=1}^{K_t} \theta_{[B_t]}^{(i)}$  where  $K_t$  is the number of scenarios. Figure 2.2b illustrates this simulation-based approach using three simulated demand scenarios, with the  $B_t$ -th largest value in each of the demand scenarios (denoted as  $\{\theta_{[B_t]}^{(i)}\}_{i=1}^3$ ) being depicted as the dashed green lines. The average of these values (depicted as a solid green line in Figure 2.2b) falls within the gap interval  $(1/4, 3/4)$ , resulting in only abilities of at least  $3/4$  being accepted at time  $t$ . The simulation-based approach yields the same action as the carefully crafted CWG policy (Algorithm 2). Interestingly, as we will later explore in Section 2.5, this simulation-based approach inherits the regret guarantee of the CWG policy (cf. Corollary 12), and outperforms the CWG policy in numerical experiments (cf. Figure 2.3b). It is worth noting that the  $\{\theta_{[B_t]}^{(i)}\}_{i=1}^3$  values represent shadow prices for the single resource (the hiring budget) under the three different demand scenarios. In the simulation-based approach, the candidate ability  $\theta_t$  is accepted if its reward  $\theta_t$  exceeds the approximated average shadow price, obtained by averaging the shadow prices over multiple demand scenarios, i.e.,  $\sum_{i=1}^3 \theta_{[B_t]}^{(i)} / 3$ . Importantly, as we present next, this simulation-based idea is not limited to the multisecretary problem but can be applied more broadly

to dynamic resource allocations, such as network revenue management and online matching, and notably inherits performance guarantees which hold for *any* algorithm satisfying certain conditions in these settings.

## 2.5 Unifying Algorithm: Repeatedly Act using Multiple Simulations

In this section, we will lift the idea of using simulations to drive decisions from the multisecretary setting to the broader class of NRM and online matching problems. We dub the resulting natural and versatile simulation-based algorithm **Repeatedly Act using Multiple Simulations (RAMS)**. Prior to formally presenting RAMS, we will establish some notations. Let  $V_t^{\text{hs}}(B_t; \theta_{\geq t})$  denote the hindsight optimal value for a given tail sequence of requests  $\theta_{\geq t} \triangleq \{\theta_t, \dots, \theta_T\}$  and remaining budget  $B_t$ ,

$$V_t^{\text{hs}}(B_t; \theta_{\geq t}) \equiv \max_{\mathbf{a} \in |\mathcal{A}|^{T-t+1}} \sum_{k=t}^T r(\theta_k, a_k) \quad \text{s.t.} \quad \sum_{k=t}^T c(\theta_k, a_k) \leq B_t. \quad (2.3)$$

Furthermore, it is natural to define  $V_{T+1}^{\text{hs}}(B_{T+1}, \emptyset) \equiv 0, \forall B_{T+1}$ . We will assume access to a simulator  $\mathcal{S}$  which takes as input a history  $\mathcal{H}$  of request arrivals and random seed  $U$  and produces a simulated demand scenario. Here a demand scenario is a tail sequence of requests  $\theta_{\geq t+1}$ ; we remark that the order of requests in a tail sequence will not matter to RAMS, since it will perform a hindsight-based calculation. Note that the assumption of access to a simulator is a weaker and more practical assumption than knowledge of the distribution  $F$ . This permits RAMS to be a data-driven algorithm where distributional knowledge  $F$  is replaced by a high fidelity simulator based on historical data. Additionally, while most of our previous discussion was focused on a stationary setting with i.i.d requests, RAMS could be applicable in non-stationary settings where the request types may have some form of temporal correlations, corresponding to the reality of many applications. This is due to the fact that RAMS is completely agnostic to the underlying type distribution.

### 2.5.1 Algorithmic Description

The basic idea behind RAMS is as follows: given the remaining budget  $B_t$  at time  $t$ , upon observing a request  $\theta_t$ , we *simulate*  $K_t$  sample paths of the future denoted as  $\{\tilde{\theta}_{\geq t+1}^{(i)}\}_{i=1}^{K_t}$ . On each of these simulated sample paths  $\tilde{\theta}_{\geq t+1}^{(i)}$ , we compute the maximum achievable cumulative reward in hindsight under each possible action  $a \in \mathcal{A}(B_t, \theta_t)$  at time  $t$ , denoted by  $Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)})$  where  $\tilde{\theta}_{\geq t}^{(i)} \triangleq \{\theta_t\} \cup \tilde{\theta}_{\geq t+1}^{(i)}$ . For each action  $a \in \mathcal{A}(B_t, \theta_t)$  we average over the  $K_t$  simulated sample paths, and choose the action which maximizes the average cumulative reward, i.e.,

$$\arg \max_{a \in \mathcal{A}(B_t, \theta_t)} K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t+1}^{(i)}).$$

We formally describe RAMS in Algorithm 3.

---

#### Algorithm 3: Repeatedly Act using Multiple Simulations (RAMS)

---

**Input:** Time Horizon  $T$ , Budget  $B \in \mathbb{R}_+^d$ , simulator  $\mathcal{S}$ , Sequence of number of simulated sample paths  $\{K_t\}_{t=1}^T$

**Initialize:**  $B_1 = B, \mathcal{H} = \emptyset$

**for**  $t = 1$  **to**  $T$  **do**

    Observe the request  $\theta_t$

$\mathcal{H} \leftarrow \mathcal{H} \cup \{\theta_t\}$

    Make  $K_t$  conditionally independent calls to the simulator  $\mathcal{S}$  with history  $\mathcal{H}$  and random seed  $U \sim \text{Unif}([0, 1])$  (denote the  $K_t$  simulated sample paths of requests as  $\{\tilde{\theta}_{\geq t+1}^{(i)}\}_{i=1}^{K_t}$ .)

**for**  $i = 1$  **to**  $K_t$  **do**

**for**  $a \in \mathcal{A}(B_t, \theta_t)$  **do**

$$Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)}) = r(\theta_t, a) + \left\{ \max_{(a_k)_{k>t}} \sum_{k>t} r(\tilde{\theta}_k^{(i)}, a_k) \text{ s.t. } \sum_{k>t} c(\tilde{\theta}_k^{(i)}, a_k) \leq B_t - c(\theta_t, a) \right\} \quad (2.4)$$

**end**

**end**

    Take the action  $a_t = \arg \max_{a \in \mathcal{A}(B_t, \theta_t)} K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)})$

$B_{t+1} \leftarrow B_t - c(\theta_t, a_t)$

**end**

---

For a feasible online policy  $\pi$ , given a state  $B_t$  and an action  $a$  which is feasible in that state

$a \in \mathcal{A}(B_t, \theta_t) \subseteq \mathcal{A}$ , define the following  $Q$ -function

$$Q_t^\pi(B_t, a; \theta_t) = r(\theta_t, a) + \mathbb{E} \left[ \sum_{k=t+1}^T r(\theta_k, a_k^\pi) \right], \quad Q_t^\star(B_t, a; \theta_t) = \max_{\pi \in \Pi(B_t - c(\theta_t, a), T-t+1)} Q_t^\pi(B_t, a; \theta_t).$$

The action under the optimal online policy is  $\arg \max_{a \in \mathcal{A}} Q_t^\star(B_t, a; \theta_t)$ , however computing this dynamic programming solution may be infeasible in general. Instead RAMS utilizes the ‘‘hindsight-based’’ approximation to the  $Q$ -function, estimated from simulated futures,  $K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)}) \approx \mathbb{E}_{\theta_{\geq t+1}} [Q_t^{\text{hs}}(B_t, a; \theta_{\geq t})]$  as a proxy to make allocation decisions. Note that  $\mathbb{E}_{\theta_{\geq t+1}} [Q_t^{\text{hs}}(B_t, a; \theta_{\geq t})] \geq Q_t^\star(B_t, a; \theta_t)$  and from (2.4), we have that  $Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)}) = V_{t+1}^{\text{hs}}(B_t - c(\theta_t, a); \tilde{\theta}_{\geq t+1}^{(i)}) + r(\theta_t, a)$ . Next we define *marginal compensation* for a given action  $a$  at time  $t$  [36]. Intuitively speaking, *marginal compensation* is the minimum payment one must make to an agent who knows the future to persuade that agent to take action  $a$  at time  $t$  on a realized sample path.

**Definition 3 (Marginal Compensation)** *Given budget  $B_t \geq \mathbf{0}$  and tail sequence of requests  $\theta_{\geq t}$  for some  $t \in [T]$ , for any action  $a \in \mathcal{A}(B_t, \theta_t)$ , we define*

$$\partial \mathcal{R}_t(B_t, a; \theta_{\geq t}) \triangleq V_t^{\text{hs}}(B_t; \theta_{\geq t}) - [V_{t+1}^{\text{hs}}(B_t - c(\theta_t, a); \theta_{\geq t+1}) + r(\theta_t, a)] \quad (2.5)$$

$$\partial \mathcal{R}_t(B_t, a) \triangleq \mathbb{E}_{\theta_{\geq t}} [\partial \mathcal{R}_t(B_t, a; \theta_{\geq t}) | B_t]. \quad (2.6)$$

We refer to  $\partial \mathcal{R}_t(B_t, a; \theta_{\geq t})$  as marginal compensation and  $\partial \mathcal{R}_t(B_t, a)$  as the expected marginal compensation. A key fact from [36, Lemma 1] is that the expected regret of a policy can be decomposed as the sum of the expected marginal compensations for the actions taken by the policy, as formalized below

**Lemma 1** *For all  $T \in [N]$  and budget  $B \in [T]$ , consider any online policy  $\pi \in \Pi(B, T)$  and let  $B_t^\pi$  denote the remaining budget at time  $t$  under policy  $\pi$ . Then we have that*

$$\text{Regret}(B, T; \pi) = \sum_{t=1}^T \mathbb{E}_{B_t^\pi} [\partial \mathcal{R}_t(B_t^\pi, a_t^\pi)]. \quad (2.7)$$

**Lemma 2 (RAMS is equivalent to minimizing expected marginal compensation)** *Given a budget  $B_t$ , request  $\theta_t$  and a collection of simulated sample paths  $\{\tilde{\theta}_{\geq t+1}^{(i)}\}_{i=1}^{K_t}$ , RAMS takes an action  $a_t \in \mathcal{A}(B_t, \theta_t)$  at time  $t$  which minimizes the simulation-based estimate of expected marginal compensation, i.e.  $a_t = \arg \min_{a \in \mathcal{A}(B_t, \theta_t)} K_t^{-1} \sum_{i=1}^{K_t} \partial \mathcal{R}_t(B_t, a; \tilde{\theta}_{\geq t}^{(i)})$  where  $\tilde{\theta}_{\geq t}^{(i)} = \{\theta_t\} \cup \tilde{\theta}_{\geq t+1}^{(i)}$ .*

Lemma 2 follows immediately from (2.4) and (2.5) and provides an alternate description of RAMS.

### 2.5.2 Performance Analysis: Meta Theorem for RAMS

Since the expected regret of the policy is the sum of the expected marginal compensations (Lemma 1), and RAMS performs a simulation-based minimization of the expected marginal compensation (Lemma 2), it follows that RAMS provides the “best achievable” regret performance (in a certain sense). This reasoning is formalized in the following meta theorem.

**Theorem 6 (Meta Performance of RAMS)** *Consider an online resource allocation problem with horizon  $T$ , number of resources  $d$ , initial budget  $B \in \mathbb{R}^d$ , a finite action set  $\mathcal{A}$  and request distribution  $F$  as defined in Section 5.2. Assume the following*

- (i) *There exists an algorithm ALG for the online resource allocation problem such that the expected marginal compensation is uniformly bounded at each  $1 \leq t \leq T$  as per  $\sup_{B_t \geq \mathbf{0}} \partial \mathcal{R}_t(B_t, a_t^{\text{ALG}}) \leq \Delta_t(\text{ALG})$  where  $B_t$  is the remaining budget at time  $t$  and  $a_t^{\text{ALG}}$  is the action under ALG.*
- (ii) *There exists a constant  $C \equiv C(F) < \infty$  such that the marginal compensation in a time step is uniformly bounded by  $C$ , i.e.,  $\sup_{B_t, a, \theta_{\geq t}} \partial \mathcal{R}_t(B_t, a; \theta_{\geq t}) \leq C$  for all  $t \geq 1$ .*

Let  $K_t$  denote the number of simulated sample paths drawn at time  $t$ . Then for any  $\eta > 2$ , there exists a constant  $C \equiv C(\eta, |\mathcal{A}|, C(F)) < \infty$ , such that

$$\text{Regret}(B, T; \text{RAMS}) \leq \sum_{t=1}^T \Delta_t(\text{ALG}) + C \sum_{t=1}^T K_t^{-\frac{1}{\eta}}$$

**Discussion of Theorem 6.** Note that while the theorem has been stated for the *i.i.d* setting, Theorem 6 can also apply to non-stationary settings with some form of temporal correlations. Theorem 6 states that the regret of RAMS can be broken down into two components:  $\Delta_t(\text{ALG})$  and  $K_t^{-\frac{1}{\eta}}$ . The former term  $\Delta_t(\text{ALG})$  follows from the assumed uniform (over the states) upper bound on the expected compensation  $\partial \mathcal{R}_t(B_t, a_t^{\text{ALG}})$  under algorithm ALG, while the latter term  $K_t^{-\frac{1}{\eta}}$  is due to the finite number of simulated sample paths. Theorem 6 states that RAMS inherits – up to sampling error – the best (uniform) regret guarantee which holds for any algorithm. Our numerical observations show that RAMS outperforms regret-optimal algorithms tailored for specific distributions or problem contexts, without the need for tuning (see Section 2.5.4). Notably, neither RAMS nor the meta theorem (Theorem 6) require prior knowledge of these optimized algorithms. As long as there exist algorithms that satisfy assumption (i) and that (ii) holds, RAMS achieves the same regret scaling.

We highlight that there exist algorithms developed in this and prior work for different problem settings which satisfy assumption (i) (cf. Corollaries 12-14). Coming to assumption (ii), in the context of network revenue management problem, this assumption holds under mild conditions, as captured in the following claim.

**Claim 1** *In the context of the NRM problem, for any request type  $\theta = (r_\theta, \mathbf{c}_\theta) \in \Theta$ , assume that the consumption vector  $\mathbf{c}_\theta$  is bounded i.e.,  $\underline{\nu} \leq \|\mathbf{c}_\theta\|_\infty \leq \bar{\nu}$  for  $0 < \underline{\nu} \leq \bar{\nu} < \infty$ . Then we have that  $\sup_{B_t, a, \theta_{\geq t}} \partial \mathcal{R}_t(B_t, a; \theta_{\geq t}) \leq d r_{\max} \bar{\nu} / \underline{\nu} \triangleq C(F)$  where  $d$  is the number of resources and  $r_{\max} \equiv \max_{\theta \in \Theta} r_\theta \leq 1$  (by assumption).*

Note that the sufficient condition in Claim 1 permits many (or infinitely many) consumption types, in contrast to the typical assumption in the prior literature of a small number of consumption types [with some notable exceptions 73, 85, 86, 70].

Combining Theorem 6 with analyses of specific algorithms, we can show that RAMS achieves the same regret scaling as that of the CWG algorithm (Algorithm 2) for the class of  $(\beta, \varepsilon_0, \delta)$ -clustered distributions (Corollary 12). Zooming out from the multisecretary problem, we consider the more general network revenue management and online matching problems. We show that under the

assumption of a small number of discrete types, RAMS achieves bounded regret scaling for both the network revenue management (Corollary 13(a)) and online matching (Corollary 14). Under infinitely many types and some structural assumptions, RAMS achieves logarithmic (Corollary 13(b)) and log-squared regret (Corollary 13(c)) scaling for the general NRM problem in line with state of the art algorithms presented in [70] and [87] respectively. Detailed assumptions and corollaries are presented in Appendix B.5 due to space constraints.

### 2.5.3 Connection of RAMS to prior work

Due to the equivalence of RAMS to minimizing the expected compensation at each time period (cf. Lemma 2), RAMS follows the ‘‘Bayes Selector’’ principle developed in [36]. However, the focus of [36] is on settings with a few types and hence their algorithm has been tailored for such settings, whereas RAMS is a very general algorithm which does not require any knowledge of the underlying assumptions on the type space.

In the context of network revenue management, RAMS is a refined version of the dual averaging policy proposed in [38], where dual prices are computed for multiple demand scenarios and the allocation decisions are made by averaging these dual prices over the different scenarios. Under RAMS, given a remaining budget  $B_t$ , a request  $\theta_t$  is accepted if  $K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, \text{accept}; \tilde{\theta}_{\geq t}^{(i)}) \geq K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, \text{reject}; \tilde{\theta}_{\geq t}^{(i)})$ . Assume that there exists a dual vector  $\mu(B_t; \tilde{\theta}_{\geq t+1}^{(i)})$  for (2.3) with tail sequence  $\tilde{\theta}_{\geq t+1}^{(i)}$  such that first order approximation of  $V_{t+1}^{\text{hs}}(B_t; \tilde{\theta}_{\geq t+1}^{(i)})$  is good, i.e.,  $V_{t+1}^{\text{hs}}(B_t; \tilde{\theta}_{\geq t+1}^{(i)}) - V_{t+1}^{\text{hs}}(B_t - c(\theta_t, \text{accept}); \tilde{\theta}_{\geq t+1}^{(i)}) \approx \mu(B_t; \tilde{\theta}_{\geq t+1}^{(i)})^\top c(\theta_t, \text{accept})$ . Then, using (2.3), (2.4) and the fact that  $r(\theta_t, \text{reject}) = 0$  and  $c(\theta_t, \text{reject}) = \mathbf{0}$ , under RAMS, the request  $\theta_t$  is accepted if

$$\begin{aligned} r(\theta_t, \text{accept}) &\geq \frac{1}{K_t} \sum_{i=1}^{K_t} \left( V_{t+1}^{\text{hs}}(B_t; \tilde{\theta}_{\geq t+1}^{(i)}) - V_{t+1}^{\text{hs}}(B_t - c(\theta_t, \text{accept}); \tilde{\theta}_{\geq t+1}^{(i)}) \right) \\ &\approx \frac{1}{K_t} \sum_{i=1}^{K_t} \mu(B_t; \tilde{\theta}_{\geq t+1}^{(i)})^\top c(\theta_t, \text{accept}) = \underbrace{\left( \frac{1}{K_t} \sum_{i=1}^{K_t} \mu(B_t; \tilde{\theta}_{\geq t+1}^{(i)}) \right)^\top}_{\text{average dual price for bid price control}} c(\theta_t, \text{accept}). \end{aligned}$$

Therefore, assuming that the first order approximation is good, RAMS will accept the request

$\theta_t$  if the reward exceeds the sum of the average dual prices for the resources it consumes, and this resembles the bid price control policy [88]. Thus we see that dual averaging is, in fact, an approximate version of RAMS for settings in which individual actions have a “small” impact, and our theoretical backing for RAMS (Theorem 6) provides new justification for why dual averaging should work well in such settings. Dual averaging is very practical and requires only a small adaptation of dual-based dynamic resource allocation systems based on model predictive control, which are typical in the industry, e.g., in supply chain optimization. Specifically, it only requires the construction of multiple demand scenarios. The hindsight problem for each scenario can be solved in parallel (using the existing MPC solver as is) and then a simple dual averaging layer can be inserted before the decision making layer.

RAMS can be viewed as the manifestation in our setting of the so-called Multi Forecast–Model Predictive Control (MF-MPC) policy which appears in the control literature, e.g., see [89] and citations therein. In MF-MPC, one constructs multiple plausible forecasts of the future, termed *scenarios*, and constructs a different plan for each of the possible scenarios, while imposing the constraint that the plans must agree on the present action to be chosen. This process is repeated each time an action is to be chosen. The connection with MF-MPC further reveals an illuminating interpretation for RAMS: Suppose all uncertainty about the future will be resolved right after the current action is chosen. What current action is optimal in this proxy problem? This is the action chosen by RAMS at each time; after all, by definition, RAMS solves the Bellman equation for this proxy problem. This interpretation throws light on the approximation underlying RAMS, and may help us –in future work– to understand how well RAMS (or, more generally, any compensation-based approach) can approximate the optimal MDP solution in a given setting.

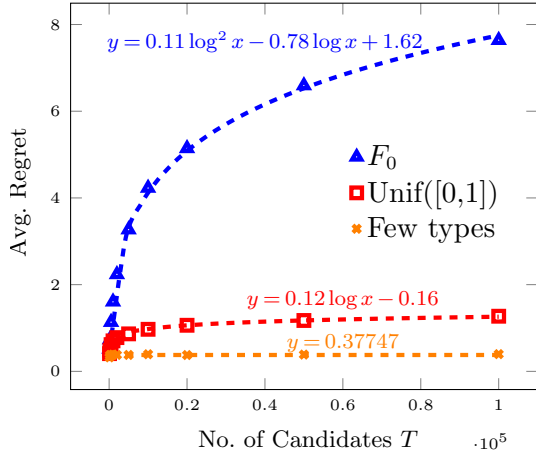
#### 2.5.4 Numerical Simulations

We perform numerical experiments under different assumptions and for different problem classes. For the multisectionary problem, we study the performance of the CWG algorithm for different distributions (Figure 2.3a), compare the performance of CE, CWG and RAMS for the

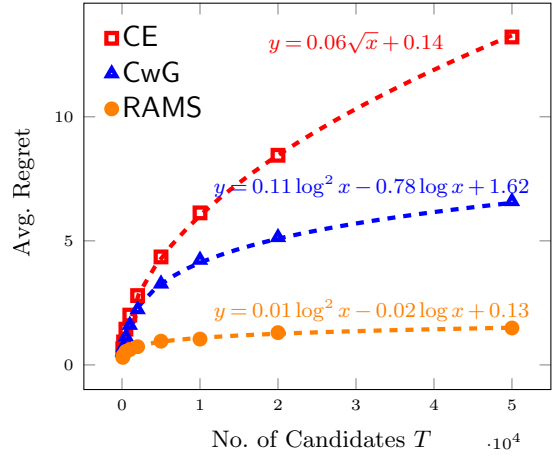
bimodal uniform distribution  $F_0$  (Figure 2.3b), and study the impact of  $\beta > 0$  (Figure 2.3c). In addition, we consider the general network revenue management problem with a few types and two resources and compare the performance of previous algorithms with that of RAMS (Figure 2.3d). In each of the settings that we consider, we vary the time horizon  $T$ , and consider a budget of  $B = T/2 \times \mathbb{1}_{d \times 1}$  where  $d$  is the number of resources. We note that this starting budget leads to the worst-case regret scaling for the instances with gaps which we consider. Overall, our simulation results confirm our theoretical predictions, including the importance of the conservativeness with respect to gaps principle, and demonstrate superior numerical performance of the RAMS algorithm.

**Figure 2.3a.** We numerically study the regret scaling of the CWG policy as a function of the time horizon  $T$  for different distributions. The distributions we consider are: (i) bimodal uniform distribution  $F_0 = \text{Uniform}([0, 1/4] \cup [3/4, 1])$ , (ii) the uniform distribution over  $[0, 1]$  and (iii) a discrete distribution over a few types  $\{0.25, 0.5, 0.75\}$  and the probability mass being  $1/3$  for each of the points. We numerically evaluate the average regret for different number of candidates  $T$  (with the budget varying as  $B = T/2$ ) and fit a curve (as shown in the dashed lines) to observe the regret scaling. For each of the three distributions considered, we empirically observe that the regret scaling is consistent with our theoretical guarantees as implied by Corollary 4 (log squared regret) for the bimodal distribution, Corollary 6 (logarithmic regret) for the uniform distribution and Corollary 5 (bounded regret) for the discrete distribution with few types.

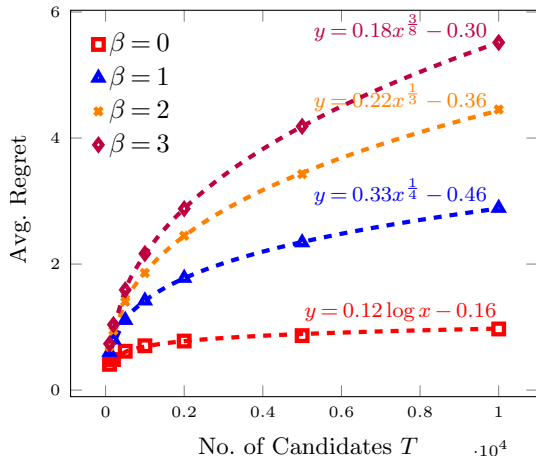
**Figure 2.3b.** We numerically study the average regret scaling of the CE, CWG and RAMS policy for the bimodal uniform distribution  $F_0 = \text{Unif}([0, 1/4] \cup [3/4, 1])$  with gap in the interval  $[1/4, 3/4]$ . We fit a curve (as shown in dashed lines) to observe the regret scaling. For each of the three policies considered, we empirically observe that the regret scaling is consistent with our theoretical guarantees as implied by Proposition 3 for the CE policy, Corollary 4 for the CWG policy and Corollary 12 for the RAMS policy. While both CWG and RAMS have the same regret scaling, we observe that RAMS has superior numerical performance over CWG since RAMS is designed to minimize the compensation and hence the regret, whereas CWG is designed to optimize



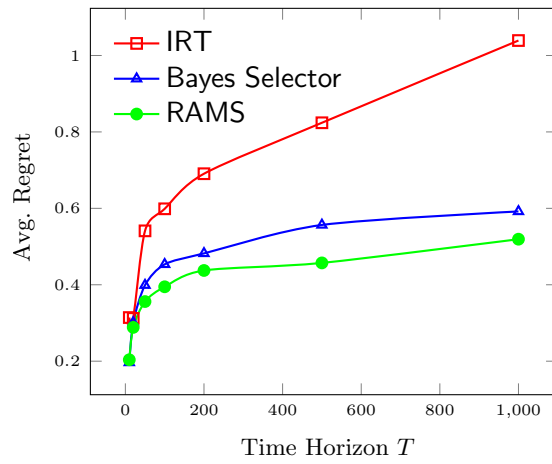
(a) CwG under different distributions



(b) Comparison of CE, CwG and RAMS on  $F_0$



(c) Polynomial regret for  $\beta > 0$



(d) Comparing IRT, BS and RAMS for NRM

Figure 2.3: (a) Illustrates the performance of CwG for different distributions, (b) compares the performance of the CE, CwG and RAMS policies on  $F_0 = \text{Unif}([0, 1/4] \cup [3/4, 1])$ , (c) highlights the polynomial regret scaling (with the exponent dependent on  $\beta$ ) for the gapless variant of the  $F_\beta$  distribution given in (2.1), (d) compares the performance of IRT, Bayes Selector and RAMS for NRM with a few types.

only the scaling of the compensation (and hence the regret scaling).

**Figure 2.3c.** To assess the influence of the parameter  $\beta$ , we examine the performance of the CE (equivalently CWG) algorithm on the *gapless* version of the  $F_\beta$  distribution, as described in (2.1) for  $\beta \in \{0, 1, 2, 3\}$ . From Theorem 4 and Corollary 6, we know that CE has the optimal regret scaling. We fit a curve (shown in dashed lines) to the empirical average regret for different values of time horizon  $T$  and observe that the regret for  $\beta \in \{1, 2, 3\}$  scales polynomially in the time horizon with the exponent given by  $\frac{1}{2} - \frac{1}{2(1+\beta)}$  and this is consistent with our guarantees in Corollary 6.

**Figure 2.3d.** We consider an NRM problem with two resources and six types. The types  $\theta = (r_\theta, \mathbf{c}_\theta)$  are given as  $\xi_1 = (1.0, [1, 0])$ ,  $\xi_2 = (0.6, [1, 0])$ ,  $\xi_3 = (1, [0, 1])$ ,  $\xi_4 = (0.5, [0, 1])$ ,  $\xi_5 = (0.9, [1, 1])$ ,  $\xi_6 = (0.8, [1, 1])$ . The requests arrive i.i.d with  $\mathbb{P}(\theta_t = \xi_j) = 0.2, \forall j \in \{1, 2, 3, 4\}$  and  $\mathbb{P}(\xi_t = \xi_j) = 0.1, \forall j \in \{5, 6\}$ . We compare the performance of RAMS against two near optimal algorithms - Infrequent Resolving with Thresholding (IRT) [37] and Bayes Selector (BS) [36]. We observe that for all the three algorithms that we consider, the regret increases initially but converges to a constant for sufficiently large  $T$ . We observe that amongst all the three algorithms considered, RAMS either matches or improves upon the algorithms.

## 2.6 Conclusion

In this work, we considered dynamic resource allocation problems and investigated the impact of distributional assumptions on algorithmic performance. By focusing on the multisecretary problem, we gained valuable insights into the fundamental drivers and limits of algorithmic regret performance. We identified a novel driver of regret, characterized by the parameter  $\beta$ , which measures the concentration of types around gaps. We introduced the Conservativeness with respect to Gaps (CwG) principle, and used it to develop an innovative algorithmic approach that mitigates the limitations of the widely used certainty-equivalent (CE) policy. The CwG principle, along with its associated CwG algorithm, achieves near-optimal regret scaling of  $\tilde{O}(T^{\frac{1}{2} - \frac{1}{2(1+\beta)}})$  for a broad class of distributions with gaps parameterized by  $\beta$ . Furthermore, we analyzed the natural Re-

peatedly Act using Multiple Simulations (RAMS) algorithm, which offers a general-purpose solution for online resource allocation problems (not just the multisectionary problem), which is applicable to any distribution of requests. RAMS is practical and data-driven, relying on simulated future demand scenarios to drive decision making. Heuristically speaking, RAMS is equivalent to a bid price control policy where the bid prices are computed by averaging the shadow prices of the hindsight optimal problem for multiple scenarios. This requires a minor adaptation of existing dual-based systems which is an industry default.

Recently, there has been a growing interest in studying online resource allocation problems in the presence of horizon uncertainty [65, 90, 67, 16]. Specifically, [67] demonstrate that by leveraging an alternative fluid benchmark, it is possible to achieve a sublinear regret scaling of  $\mathcal{O}(\sqrt{T})$ , through the use of a static policy. Nevertheless, a naïve implementation of the RAMS approach yields regret (relative to the alternative fluid benchmark of [67]) that scales linearly. Whether RAMS can be adapted to attain sublinear regret remains unknown. We leave the exploration of this intriguing question, as well as other related queries surrounding the development of near-optimal algorithms under horizon uncertainty, for future endeavors.

## Chapter 3: MOTIF: Multi-Objective Tradeoff in Fulfillment

### 3.1 Introduction

We collaborate with a large scale online retailer to optimize their order fulfillment pipeline. For each one of the millions of orders that the online retailer processes daily [91], the order assignment engine (OAE) is responsible for several discrete decisions to ensure demand gets fulfilled within the promised timeframe. These decisions include: *(i)* what items are boxed together to form a shipment, *(ii)* which warehouse will process each shipment, *(iii)* whether we need to transfer items between warehouses for consolidation, *(iv)* which transportation route to use for each shipment, including whether we can synchronize deliveries to the customer destination and, critically, *(v)* when each of these series of operations will take place?

At the core, the order assignment system employs a mixed integer programming model to make its order assignment decision. The model is a variant of the set-partitioning problem [92] initialized for each order with a discrete set of candidate shipments. A candidate shipment represents one possible way to assign an order in full or in part, determining what items to pick for the shipment, the origin warehouse, as well as transportation route and timings. While for small orders one can exhaustively enumerate all candidate shipments, OAE leverages local search heuristics to select a heterogeneous subset of candidate shipments for larger orders. The model is sufficiently small that it can be solved in less than a second for majority of orders, a strict requirement to ensure we reserve inventory and transportation capacity for each demand. Naturally, order assignment decisions are subject to re-evaluation at any point, including periodic re-optimization of all pending orders simultaneously to resolve inefficiencies of individual order decisions.

Historically, objectives relevant to network efficiency and operating costs (e.g., unit-per-box (UPB), in-region assignment (IRA), reactive transfers) were a consequence of the complex in-

teractions of shipment level cost components, which were the sole guide to OAE. However, OAE realized that shipment level cost components alone were insufficient to exert direct control on some of the network objectives. This led to the Cost Relaxation (CR) approach. OAE leveraged the Cost Relaxation (CR) approach to reduce the dependency on shipment level cost components and instead explicitly optimize for network level objectives.

However, CR is rife with several limitations. Firstly, CR relies on a predefined hierarchy of objective importance, and the objectives ranked lower in the hierarchical order (e.g., non-monetary costs) can be entirely ignored for gains on other objectives, irrespective of the gain amount. Simply put, there is no notion of tradeoff between objectives. Stacking objectives in a strict cascade is inadequate because specific strategic outcomes could become more or less important as our network conditions change. Secondly and more importantly, CR produces Pareto inefficient solutions at the macro-level, i.e., CR leads to order assignments that can be further improved along one or more objectives without negatively affecting the others. The underlying reason is that CR does not trade off consistently between objectives across different demand sets, as discussed in Section 3.2. Finally, the CR approach is convoluted as it starts from a human-judgement call, iterating configuration value manually through rounds of simulations and then observing how much results deviate in production. Its complexity has increased vastly as OAE kept adding layers to the objective list.

To address the aforementioned challenges, we completely overhauled the order assignment logic by introducing a new optimization model, Multi-Objective Tradeoff in OAE (MOTIF). MOTIF optimizes a *single* blended objective (BO), i.e., a single weighted linear combination of the multiple objectives considered by CR, and it is guaranteed to produce (approximately) Pareto efficient solutions at an aggregate level provided the underlying problem satisfies certain conditions. MOTIF's design is grounded in the idea that while at the micro-level (per order level), the optimization problem is nonconvex, at the macro-level (across millions of orders) the problem is approximately convex (due to Shapley-Folkman Theorem [1, 2]) – this necessitated a critical mindset shift from a combinatorial viewpoint to a convexity viewpoint. In addition to this foundational different approach, we make additional engineering innovations to operationalize and scale MO-

TIF: (i) near real time generation of Pareto frontiers using the augmented  $\varepsilon$ -constraint method to enable business leaders to choose their preferred operating regimes under different network conditions and (ii) symmetry reduction techniques to reduce the number of decision variables, minimize solve time and reduce millions of MIP solves per day.

There are two key challenges in operationalizing MOTIF: (i) the choice of weights for the blended objective which align well with certain business objectives; (ii) a selection of the targeted values for the business objectives under different network conditions. We decouple the two challenges by addressing (i) with a principled approach to generating weights, resulting in solutions which improve upon CR solutions; and tackle (ii) with an algorithm which characterizes the Pareto frontier in the objective space. These two capabilities together enable building a fast decision support tool allowing business leaders to specify the business goals in global terms (i.e., the operating target on the frontier), and the weight generation algorithm provides weights aligning with these business goals in near real time.

The blended objective approach made the tradeoffs between different objectives consistent and enabled better control, allowing OAE to adapt to various network conditions. The BO solution resulted in Pareto improvements over CR. After two large scale experiments with MOTIF on a subset of regions in the United States, the model has been fully rolled out in **entire** North America (NA) and is responsible for making order assignment decisions for millions of orders per day.

MOTIF ..

1. improved in-region assignments and consolidation (units per box) while reducing cost per unit.
2. significantly reduced computational overheads and scaled seamlessly with network growth thanks to the modular approach.

## 3.2 Pitfalls of Cost Relaxation (CR)

### 3.2.1 Inconsistent Objective Tradeoff in CR leads to Pareto Inefficiency

We briefly outline the CR approach and its pitfalls via a stylized illustration. Suppose there are two objectives: `cost` (measured in dollars) and `num_packages` (number of packages). The objective `num_packages` matters since the last mile cost, which is outside of the first `cost` objective calculation, is roughly proportional to the package count. Under CR, one first decides the prioritization between the objectives, e.g., cost reduction as the primary objective, with reducing the package count being the secondary objective. CR then proceeds for each order by first finding the cost minimizing assignment, and then looking for the assignment that minimizes the number of packages, subject to the constraint that cost should not exceed the minimum cost by more than  $\alpha\%$ , where the permitted relaxation  $\alpha$  is a tuning parameter of CR. It turns out that CR described above, with  $\alpha\%$  relaxations in higher priority objectives suffers from a fundamental drawback, namely, it trades off inconsistently between the two objectives. Consider the setup above and two orders, A and B. The number of items of each order and the objective values of their available fulfillment options are summarized in Table 3.1.

Table 3.1: Demand and Fulfillment Options

Order index	#items	Option 1	Option 2
A	3	(\$5, 3 packages)	(\$8, 2 packages)
B	2	(\$2.5, 2 packages)	(\$4.5, 1 package)

Suppose each additional package is estimated to incur \$2.5 in last mile costs on top of the cost specified in the fulfillment options. Then the optimal solution assigns order A to Option 1 (its total cost of  $\$5 + 3 \times \$2.5 = \$12.5$  is 50¢ cheaper than the alternative option); and order B to Option 2 (its total cost of  $\$4.5 + \$2.5 = \$7$  is also 50¢ cheaper than the alternative option). CR, however, fails to assign both orders correctly irrespective of the value of  $x$ . For example, if we allow a  $\alpha = 70\%$  cost increase, it picks Option 2 for order A since the cost increase is only 60%, and Option 1 for order B, since the cost increase under the alternate assignment would have been 80% which is

larger than permitted. It assigns *both* orders sub-optimally. More broadly, for all  $\alpha$ , it assigns at least one order sub-optimally: it assigns order A sub-optimally for all  $\alpha \geq 60$  above and order B sub-optimally for all  $\alpha \leq 80$ . This inconsistency at the micro-level leads to Pareto inefficiencies at the macro-level. We did extensive numerical investigations on real data to show that CR resulted in Pareto inefficient solutions.

The solution to this inefficiency of CR is to tradeoff consistently between the two objectives based on an appropriately chosen “conversion rate” between them. In this case, the optimal solution can be recovered by assigning each order to the option which minimizes the single blended objective  $\$cost + \$2.5 \times num\_packages$ . This is precisely what MOTIF achieves using the BO approach: make the trade-offs consistent and achieve (approximate) Pareto efficiency at the macro level.

### 3.2.2 Increased Complexity and Computational Overheads

The original version of CR consisted of a single relaxation solve - up to  $\alpha\%$  of total non-monetary costs for each shipment reduced. However, over time OAE stacked a number of layers on top of the original CR framework to solve for regionalization and transship reduction (referred as CTR<sup>1</sup> logic). These required additional relaxation layers, making CR increasingly complicated and unwieldy. Note that as one adds layers to CR, the additional opportunities to change the solution reduces in the subsequent layer, limiting ones’ capability to achieve intended goal with newly added objective. Moreover, CR logic changes require an intensive hands-on-the-wheel process to evaluate outcomes of different configurations, with substantial efforts in engineering to amend the CR model with additional logic, set up simulations and production experiments to estimate impact, validate code performance and iterate on the options considered for final decision. With MOTIF, we are largely simplifying the model specification in OAE, moving away from 10+ MIP solves with CR to a single MIP solve.

---

<sup>1</sup>“C” stands for Consolidation (shipment reduction), “T” for Transship, “R” for in-Region assignment.

### 3.3 Blended Objective (BO) Approach: Theoretical Foundations

#### 3.3.1 Notation and Definition

Let  $\mathcal{O}$  be the set of objectives (metrics) that OAE optimizes for. This includes monetary cost, non-monetary cost, UPB (units per box), IRA (in region assignments), and transship usage. Let  $F = (\mathcal{S}, \mathbf{d})$  denote a single FSet<sup>2</sup> with candidate shipments set  $\mathcal{S}$  and a demand vector  $\mathbf{d}$ . Each candidate shipment  $s \in \mathcal{S}$  is associated with the objective scores which are denoted as  $c_s^k$  for  $k \in \mathcal{O}$ . Furthermore, let  $\mathbf{P} \in \mathbb{Z}^{|\mathcal{d}| \times |\mathcal{S}|}$  denote the matrix of which each column indicates the quantity of items each candidate shipment contains. Finally, the decision for the order assignment problem are to select a subset of candidate shipments to fulfill the demand, i.e.,  $x_s$  is the binary variable indicating whether shipment  $s$  is used to fulfill demand for all  $s \in \mathcal{S}$ . We define the set of feasible solutions as  $\mathcal{X}_F = \{\mathbf{x} : \mathbf{P}\mathbf{x} = \mathbf{d} \text{ and } \mathbf{x} \in \{0, 1\}^{|\mathcal{S}|}\}$ , also referred as the solution space. We assume that  $\mathcal{X}_F$  is always non-empty, the feasibility is guaranteed by the candidate shipment generation process. Given a feasible solution  $\mathbf{x} \in \mathcal{X}_F$ , we define the  $k$ -th objective value achieved by the selected candidate shipments as follows:

$$f_F^k(\mathbf{x}) = \sum_{s \in \mathcal{S}} c_s^k x_s, \quad \forall k \in \mathcal{O}. \quad (3.1)$$

At the individual FSet level, the goal of OAE is to minimize a multi-objective optimization problem:

$$\min_{\mathbf{x} \in \mathcal{X}_F} \left\{ \mathbf{f}_F(\mathbf{x}) \triangleq \left( f_F^k(\mathbf{x}), k \in \mathcal{O} \right) \right\} \quad (3.2)$$

Let  $\mathcal{F}$  denote a collection of individual FSets. We refer to the solution space corresponding to a single FSet as the *micro level decision space* and the solution space corresponding to a collection  $\mathcal{F}$  as the *macro level decision space*. Let  $\pi$  denote a procedure used to solve the problem in (3.2),  $\mathbf{x}_F^\pi$  denote the solution corresponding to FSet  $F$ . We call  $\mathbf{f}_F^\pi \triangleq \mathbf{f}_F(\mathbf{x}_F^\pi)$  the micro multi-objective values (MOV) attained by solution  $\mathbf{x}_F^\pi$  for the single FSet. We will only focus on the objective

---

<sup>2</sup>An FSet, or fulfillment set, contains a set of items from the same customer, possibly from different orders.

values at the *macro* level. Therefore, we define  $\mathbf{f}_{\mathcal{F}}^{\pi} = \sum_{F \in \mathcal{F}} \mathbf{f}_F^{\pi}$  as the aggregated (macro) MOV for the collection of FSets  $\mathcal{F}$ . We call a macro MOV *Pareto efficient* if it is impossible to improve a particular metric while not making any other metric worse off. We refer to a macro MOV as a *CR dominating* solution if it is as good as that produced by the CR procedure along all metrics and strictly better than CR on at least one metric.

### 3.3.2 Blended Objective (BO) Formulation

Instead of solving for the different objectives in a cascade, the BO approach solves a single optimization problem which is a linear combination of all. In particular, for a given FSet  $F$  and a given set of weights  $\lambda \in \mathbb{R}_{\geq 0}^{|\mathcal{O}|}$ , the BO approach is to solve the following integer program.

$$\min \sum_{k \in \mathcal{O}} \lambda^k f_F^k(\mathbf{x}) \quad (\text{blended objective}) \quad \text{subject to } \mathbf{x} \in \mathcal{X}_F \quad (3.3)$$

Apart from its simplicity, the BO approach also addresses the Pareto inefficiency issue in the CR approach. To see this point, we need to formulate the BO problem in the macro solution space.

$$\min \sum_{F \in \mathcal{F}} \sum_{k \in \mathcal{O}} \lambda^k f_F^k(\mathbf{x}_F) \quad \text{subject to } \mathbf{x}_F \in \mathcal{X}_F \quad \forall F \in \mathcal{F}. \quad (3.4)$$

While the individual FSet problem (3.3) is non-convex, the Shapley-Folkman theorem guarantees that the macro problem (3.4) is approximately convex on a large enough collection of FSets, provided that the contribution of each micro decision to the macro MOV is small, and the number of constraints is much smaller than the number of micro decisions [2, 1] — both conditions apply to the OAE use case. As a result of the macro optimization problem (3.4) being approximately convex, the BO approach will lead to approximately Pareto efficient solutions with non-negative weights [93].

### 3.4 Principled Framework for obtaining CR-dominating solutions

Given CR is Pareto inefficiency, we want to generate a set of weights resulting in a macro MOV that dominates the CR macro MOV, i.e.,  $\mathbf{f}_F^{\text{BO}} \leq \mathbf{f}_F^{\text{CR}}$ , but with at least one dimension with strict inequality holds. Given a collection of historical FSets  $\mathcal{F}$ , let  $\mathbf{v}_{\text{CR}} \in \mathbb{R}_{\geq 0}^{|\mathcal{O}|}$  be the macro MOV obtained by the CR solution across these FSets. We further simplify the notation by defining  $\mathcal{X}_F = \{\mathbf{x} = (\mathbf{x}_F)_{F \in \mathcal{F}} : \mathbf{x}_F \in \mathcal{X}_F\}$ , and  $f^k(\mathbf{x}) = \sum_{F \in \mathcal{F}} f_F^k(\mathbf{x}_F)$  as the macro objective value for the  $k$ -th objective aggregated over  $\mathcal{F}$ . Without loss of generality, we use monetary cost as the main objective (i.e., `obj = monetary cost`) in the multi-fset formulation (3.5) below, and it can be switched with any other objective in the objective set  $\mathcal{O}$ .

$$\min_{\mathbf{x} \in \mathcal{X}_F} f^{\text{obj}}(\mathbf{x}) \quad (\text{minimize one aggregated objective value across FSets}) \quad (3.5a)$$

$$\text{subject to } f^k(\mathbf{x}) \leq v_{\text{CR}}^k \quad \forall k \in \mathcal{O}' \triangleq \mathcal{O} \setminus \{\text{obj}\} \quad (\text{being no worse than CR on other metrics}) \quad (3.5b)$$

Note that the partial dual optimization problem for (3.5) is given as

$$\max_{\lambda \in \mathbb{R}_{\geq 0}^{|\mathcal{O}'|}} \min_{\mathbf{x} \in \mathcal{X}_F} \left( f^{\text{obj}}(\mathbf{x}) + \sum_{k \in \mathcal{O}'} \lambda^k f^k(\mathbf{x}) \right) - \sum_{k \in \mathcal{O}'} \lambda^k v_{\text{CR}}^k \quad (3.6)$$

Observing (3.6), the weights for the BO approach are just the duals corresponding to the constraints in (3.5b) and hence the problem essentially reduces to finding the optimal dual solution corresponding to (3.5b). Due to the Shapley-Folkman Theorem, the primal problem (3.5) is approximately convex, and the duality gap is small given the number of constraints is much smaller than the number of micro decisions and it vanishes as the number of FSets grows to infinity [1, 2]. One can leverage primal dual algorithms to solve for the dual values iteratively [94]. Empirically, we also observed that the integrality gap of problem (3.5) is very small. For problems with roughly 5000 FSets, the gaps is already less than 0.1%. This permits the use of LP relaxation dual as the

weights for the BO solve. Note that  $\lambda_{\text{obj}} = 1$ , i.e. a normalization of the dual vector, in the final weights.

One major issue with this approach is that typically there are infinitely many optimal dual solutions to for (3.5), all of which could land a CR dominating solutions. However, some of these solutions do not perform consistently for out of sample FSets failing to deliver similar macro objective values as in the training data. For this reason, instead of using the Simplex method, we use the Barrier algorithm without Crossover to generate the weights as duals. The duals generated from the Barrier algorithm perform more robustly on out of sample FSets. The dual problem admits multiple solutions and the Barrier algorithm tends to average these out as opposed to the Simplex method yielding a dual solution that is an extreme point.

### **3.5 Pareto Frontier: A decision support tool for business leaders**

Multi-objective optimization rarely has a single best solution that optimizes all objectives. Instead, a set of solutions exists where improving one objective requires compromising another, known as the Pareto frontier. A solution on this frontier is *Pareto efficient*, meaning no objective can be improved without harming another. For example, in an order fulfillment scenario, if a solution is already Pareto efficient, lowering shipping costs would require either more split shipments or a lower IRA. Understanding this frontier is crucial because it reveals the limits of what can be achieved when juggling multiple objectives. The Pareto frontier is an essential tool for decision-making in complex, multi-objective settings. By mapping this frontier, decision-makers can visualize and quantify the tradeoffs between objectives, allowing them to choose acceptable compromises based on current business priorities and constraints. Instead of relying on guesses or trial and error, a clear picture of the frontier supports systematic exploration of different operating points and their associated tradeoffs. This becomes especially important when business conditions or priorities change, as the frontier shows the full range of achievable outcomes and their costs. Consequently, a systematic approach to profiling the Pareto frontier in the macro objective space is needed.

We implemented the augmented  $\varepsilon$ -constraint method (also referred to as AUGMECON) to profile the Pareto frontier in the multi-objective space in OAE [95]. The algorithm works as follows: given a grid points across all objectives, it repeatedly solves an optimization problem for each grid point, and finds a Pareto efficient solution close to the grid point. The algorithm employs a mechanism to cut off certain grid section based on information gathered from solves completed. We first determine the range for each objective by solving lexicographic goal programming with different objective ordering. Next, we divide the range into equidistant grid points to obtain the desirable RHS values  $\mathbf{v}$  for the constraints. By varying the RHS values, the algorithm returns a set of Pareto efficient MOVs. It is worth noting that we can generate a more granular view of the frontier by customizing the grid, for instance, only focusing on the area that dominates the CR solution. A byproduct of these individual solves is the proper weights to achieve the corresponding efficient MOVs.

When determining the number of FSets included in the multi-FSet problem, we face a tradeoff between the accuracy of the frontier and the computational burden of AUGMECON. To address this problem, we implemented a symmetry reduction<sup>3</sup> method to reduce the redundancy of solutions space of problem hence making it easier to solve.

Specifically, we define an *equivalence relation* that captures the symmetry in the problem. In MOTIF an equivalence class is a class that contains all candidate shipments which contribute the same amount to the solution quality for each of the following metrics: line item ID, quantity of line item, cost, transshipped units, in-region fulfillment, and direct lane assignment. If there are multiple candidate shipments with the same numerical values for all these metrics, then they form a group, and a only a single variable is created to represent the group. However, we must assign an upper bound to this variable equal to the number of candidate shipments in the group. Our experiment results demonstrate 5-10 $\times$  reduction in the number of decision variables and a corresponding reduction to run time at no impact to quality of results.

---

<sup>3</sup>Symmetry reduction in Integer Linear Programming (ILP) refers to techniques used to reduce the number of symmetric solutions in the problem's solution space. Symmetries occur when multiple solutions to an ILP problem are essentially equivalent due to the problem's structure. Symmetries cause redundant calculations and increase the complexity of finding an optimal solution.

The Pareto frontier could change over time based on various factors, such as network condition, time of the year (Peak vs Business-As-Usual), etc. We periodically refresh the frontier using latest FSet samples and there is active research work on the robustness of the frontier.

### **3.6 Wide-scale Rollout and Performance**

Through two large scale pilots, MOTIF demonstrated consistent Pareto dominance over the CTR logic. The success of these pilots convinced the leadership to green light a wide-scale rollout of MOTIF. Subsequently, we rolled out MOTIF to the **entire** U.S. network. Since then, we have deployed three different weight configurations to handle high-volume events. The first set of weights outperformed the counterfactual CR solution by improving UPB, raising IRA and reducing monetary cost-per-unit (CPU) and reactive transship usage. It also boosted local assignment (LA), direct-lane assignment (DLA), and units per paid delivery (UPPD). Subsequently, we introduced another set of weights to further optimize local assignments (LA) and direct-lane assignment (DLA) while retaining Pareto dominance over CR for key objectives. Compared to the previous set of weights, we further increased the local assignments (LA) and direct-lane assignments (DLA) at the cost of increased reactive transships. As peak approached, in order to keep up with the high anticipated demand and in accordance with the business needs of the time, we introduced another set of weights which improved in region assignments (IRA), local assignments (LA) and direct-lane assignments (DLA) while increasing reactive transship usage.

## Chapter 4: The Fault in Our Recommendations: On the Perils of Optimizing the Measurable

*Based on the paper [96] co-authored with Omar Besbes and Yash Kanoria.*

### 4.1 Introduction

Recommendation systems have become integral to numerous online platforms, including social media, e-commerce sites, and media streaming services. These algorithmic mechanisms are deployed by digital platforms to suggest content, products, and services potentially appealing to users. A key promise of these systems is their ability to unveil *hidden gems*—content, products, or services users might not have known about but would derive immense *utility* from. However, there are questions regarding recommendation systems ability to enable sufficient exploration and the discovery of *niche* items which may potentially be of high value to some users. Existing literature [97] and anecdotal evidence suggests a tendency towards a “popularity bias” in contemporary recommendation algorithms, where the algorithms are skewed towards suggesting items that are already popular. The heart of this problem seems to lie in the signals that are measured and optimized for by these recommendation systems. Most recommendations systems use a ranking-based logic to sort and display content by predicted engagement [4, 98]. The signals used to predict and optimize for engagement are generally clicks, likes, comments, and sometimes also include continuous measures of interactions such as dwell time and watch time. While these signals may serve as proxies for *utility*, there is potentially a large misalignment between these signals and the *utility* generated for the user by the recommended items [3]. The fact that a user clicks on a piece of content does not tell us much about how much they *value* the content. There has been little concerted effort to measure the *user utility* generated by the recommended items. The reason

for this is that *utility* is quite challenging to measure, and most recommendation systems optimize for signals which can be directly measured, such as clicks, purchases or consumption times. This motivates our key research questions.

*What are the implications of engagement maximization on user utility? Can one improve user utility without measuring it explicitly? And if so, what are the implications on engagement?*

The broader question of misalignment between engagement and utility maximization by recommendation systems has recently received attention in several papers [99, 4, 100]; largely through case studies or empirical investigations. We contribute to this growing line of work by studying a stylized model of recommendation systems, allowing us to develop structural insights about the aforementioned misalignment and how it may be addressed. We study a parsimonious model of a multi-item recommendation system and a repeated content consumption model, where at each interaction, the user selects between an outside option and the best option from a recommended set of items. The recommendation system decides the type(s) of content to recommend at each iteration given a constraint on the number of items recommended, and optimizes for a specified objective. We assume that there are only two types of items – a popular type (P) which has a fixed positive mean utility and a niche type (N), the utility of which is distributed according to some distribution  $F_N$  and has a mean utility of zero. Additionally, the user has an outside option which we assume has a mean utility of zero as well. This modelling choice allows us to capture the key tension in recommendation systems – popular items have similar utility across users whereas niche items can have a high variance across users in terms of their utility. Niche items are not every users’ “cup of tea” and generally there is a huge variation in the utility generated by the niche items. We model niche items as being of low utility for a large fraction of users and being of high utility for a small fraction of users, such that on average the utility is assumed to be zero. In particular, *on average across users* the popular type provides more utility than the niche type. For simplicity of exposition, we assume that the platform recommends two items at each iteration. We now summarize the main contributions of our work.

(i) *Theoretical analysis under stylized assumptions.* In Section 4.3, we will assume that the plat-

form has prior knowledge about the distribution of type utilities, which will be subsequently relaxed in Section 4.4. Furthermore, we will assume that the utility of the popular type is fixed and known and the utility of the niche type is a two-point distribution with a single parameter  $p$  (cf. (4.5)). The parameter  $p$  is the fraction of users who derive high utility from the niche type. Our analysis in Section 4.3 will focus on the regime where the parameter  $p$  is small.

(a) *Stark structural misalignment between engagement and utility maximizing policies.*

Our analysis reveals a key structure misalignment between engagement maximizing policies and utility maximizing policies. We show that in the regime of interest ( $p \rightarrow 0$ ), it is engagement maximizing to never recommend the niche types, i.e., the engagement maximizing policy results in homogeneous recommendations, where, at each iteration, it only recommends items of the popular type (cf. Theorem 7). This result, while extreme, is indicative of insufficient exploration and inadequate diversity in many modern day recommendation systems. In contrast, we develop and study a simple utility-aware heuristic PEAR (Algorithm 4) which recommends a diverse set of items, and attains significantly higher user utility than the engagement maximizing policy (compare Theorems 7 and 8).

(b) *The best of both worlds.* Apart from illuminating the existence of a potentially stark misalignment between engagement and utility, our analysis also characterizes the magnitude of the aforementioned misalignment. Quite strikingly, our analysis uncovers asymmetry in the misalignment between engagement and utility. We observe that by optimizing solely for engagement, there can be substantial loss in the (expected) user utility irrespective of how forward-looking the platform is. In contrast, our utility-aware heuristic PEAR (Algorithm 4) not only achieves approximately optimal utility (Corollary 7) but also near optimal engagement (Corollary 8), as the platform becomes increasingly forward looking. Corollary 8 and Figure 4.1 highlight this asymmetry where we observe a sharp decrease in utility for a minuscule gain in engagement as we move

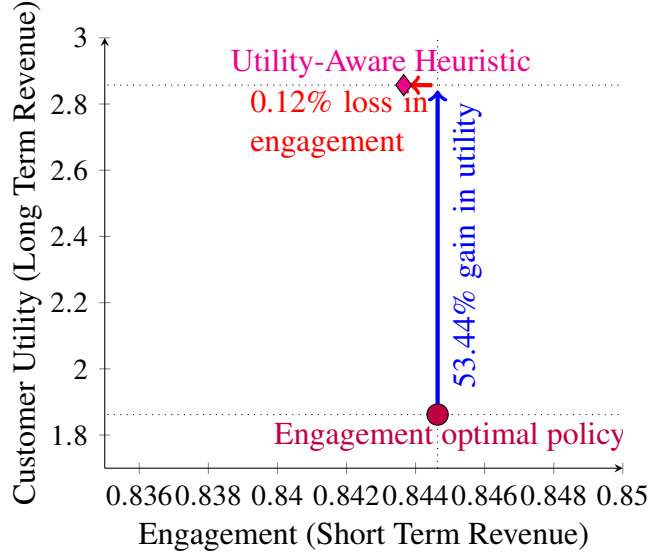


Figure 4.1: A comparison of different policies showing that it is possible to improve utility substantially from an engagement-maximizing policy with minimal loss in engagement. Note that the discount factor is  $\delta = 0.99$ .

from the utility-aware heuristic PEAR to the engagement optimal policy APP (defined in (4.6)). These findings suggest that platforms are significantly under-delivering on *user utility* by optimizing solely for engagement, due to inadequate exploration. Table 4.1 highlights the magnitude of the misalignment as a function of how forward looking the platform is which is captured by the discount factor  $\delta \in [0, 1)$ . (Small values of  $\delta$  correspond to platforms being myopic.) In Table 4.1, the mean utility of the popular type is assumed to be  $V_P = 1$ . In the first and second row of the table, we have the relative changes in engagement and utility under the utility-aware heuristic PEAR with respect to the engagement optimal policy APP, respectively.

Table 4.1: Loss in engagement and gain in utility of PEAR compared to the engagement-maximizing policy APP

	$\delta = 0$	$\delta = 0.9$	$\delta = 0.99$	$\delta = 0.999$
$\Delta$ Engagement	-10.6%	-1.2%	-0.12%	-0.011%
$\Delta$ Utility	+29.3%	+51.1%	+53.4%	+53.7%

(ii) *Numerical Evidence of generalizability of insights.* In Section 4.4, we relax the assumptions

that the platform *knows* the mean utility of the popular type or the distribution of utilities for the niche type, which is in line with practice where data on possible *user utility* values is seldom available. Moreover, instead of assuming a two-point distribution for the utility of the niche type, we consider the class of Pareto distributions with different scale parameters [101]. Since it is technically challenging to characterize the engagement or welfare optimal policy under the general settings considered above, we will restrict our attention to specific policies – (i) **Always Popular Policy (APP)** (see (4.6)), which as the name suggests, only recommends items of the popular type and (ii) **Dlverse-then-CustomizEd (DICE)** policy (Algorithm 5), which recommends a mix of popular and niche type of items for the initial  $\mathcal{T}$  periods and switches to recommending either popular or niche type depending on the user’s choices in the initial  $\mathcal{T}$  periods. Note that neither policy measures or estimates the *user utility* from the recommendations. Through a numerical analysis, we observe that as the tail of the Pareto distribution becomes *lighter*, the no exploration policy **APP** does strictly better than the exploration driven heuristic **DICE** on engagement, but by a minuscule amount. On the flip side, there is significant utility loss incurred by **APP** vis-a-vis the **DICE** heuristic (cf. Figure 4.4). These results highlight that one can improve substantially upon the engagement maximizing **APP** policy in terms of *user utility* – despite not being able to measure it – by making diverse and exploratory recommendations, without substantially reducing engagement.

#### 4.1.1 Related Literature

*Popularity Bias in Recommendation Systems.* Many present-day recommendation systems are plagued by the “popularity bias” where algorithms disproportionately favor already popular items, thereby neglecting niche content that could be valuable to users [102, 97, 103, 104]. Recommendation systems provide users with a list of items typically ranked in a descending order of predicted engagement [98, 105], and the popularity bias is often attributed to this ranking logic [106]. The topic of popularity bias has also become relevant from a fairness and bias in recommendation

systems point of view [107, 108].

*Measuring Value Beyond Engagement.* The need for more refined measures of engagement has been highlighted in various contexts [109, 110, 111]. A critical challenge in recommendation systems has been measuring the true *utility* of recommendations to the users, which extend beyond mere engagement metrics such as clicks and dwell times. Many industrial grade recommendation systems try to incorporate different non-engagement signals as a proxy for utility [98, 112]. Recent work uses a measurement theory approach that constructs a more comprehensive view of value, integrating both observed engagement signals and latent variables to better capture user satisfaction [4], however the focus of [4] is on determining *whether* a user values certain content rather than *how much* they value it.

*Diversity and Novelty in Recommendation Systems.* Instead of solely optimizing for engagement, many recommendations systems have started optimizing for other metrics such as diversity and novelty (refer to [113] for the definition of diversity and novelty in recommendation systems) and take a multi-objective optimization approach [112]. There is a growing consensus that diverse recommendations are valuable for users [114, 115, 116] and they help in improving long term retention on platforms [99]. Diverse recommendations also help in covering a user’s diverse set of interests and mitigate the saturation effects resulting from consuming homogeneous content [117]. Novelty is another related concept which has been studied in recommendation systems from the perspective of discovery of niche content [118] and this is intimately related to the concept of long tail recommendations [119].

## 4.2 Model

We consider an infinite horizon setting with a discount factor  $\delta \in [0, 1)$ . At each time  $t \in \mathbb{N}$ , the recommendation system recommends  $K$  items. For ease of exposition, in this paper we will focus on  $K = 2$ . There are two *types* of items which we refer to as popular (**P**) and niche (**N**) types. There are infinitely many items corresponding to each of these two types. We denote the set of items corresponding to the popular and niche types as  $\mathcal{I}_P$  and  $\mathcal{I}_N$  respectively and let  $\mathcal{I} = \mathcal{I}_P \cup \mathcal{I}_N$

be the set of all the items. For a given item  $i \in \mathcal{I}$ , let  $\tau : \mathcal{I} \rightarrow \{P, N\}$  denote the type of the item. We denote the set of recommended items at time  $t$  as  $\pi_t = \{i_{1,t}, i_{2,t}, \dots, i_{K,t}\}$  where  $i_{j,t}$  refers to the  $j$ -th item recommended at time  $t$ . We denote  $\pi = (\pi_0, \pi_1, \pi_2, \dots)$  as the recommendation policy of the platform. The recommendation policy first decides on the type of item to recommend either P or N and then recommends an item which has not been consumed by the user previously from either  $\mathcal{I}_P$  or  $\mathcal{I}_N$ . Given a set of recommended items  $\pi_t$  at time  $t$ , the user chooses at most one item using an underlying choice model. Let  $c_t \triangleq c_t(\pi_t)$  denote the item chosen by the user. Note that we allow for the user to not choose any of the recommended options, in which case we denote the chosen item as  $\emptyset$  and refer to  $\emptyset$  as the outside option. Each of the two item types, popular and niche, has a base utility, denoted as  $V_P$  and  $V_N$  respectively. We assume that the base utility corresponding to the popular product  $V_P$  is fixed, whereas the base utility corresponding to the niche product  $V_N$  is a random variable with distribution  $F_N$ . We adopt the classical multinomial logit choice model studied in discrete choice literature [120]: The utility (or value) that the user derives by consuming item  $i$  of type  $\tau(i)$  is given as

$$u_i = V_{\tau(i)} + \epsilon_i, \quad (4.1)$$

where  $\epsilon_i$  is an idiosyncratic noise term which is assumed to be a standard Gumbel distribution, i.e.  $F_\epsilon(x) = \exp(-\exp(-(x-\gamma)))$ , where  $\gamma$  is the Euler-Mascheroni constant. The noise term  $\epsilon$  is zero mean, i.e.  $\mathbb{E}[\epsilon] = 0$ . The user is a utility maximizer, i.e., given a set of  $K$  recommended items  $\pi = \{i_1, i_2, \dots, i_K\}$ , the user chooses the item with the highest utility

$$c(\pi) = \arg \max_{j \in \pi \cup \{\emptyset\}} u_j = \arg \max_{j \in \pi \cup \{\emptyset\}} V_{\tau(j)} + \epsilon_j, \quad (4.2)$$

where  $\epsilon_j$  are independent draws from the common distribution  $F_\epsilon$ . We assume that  $V_{\tau(\emptyset)} = 0$ , i.e.,  $u_\emptyset \sim F_\epsilon$  and we denote  $\tau(\emptyset) \triangleq \mathbf{O}$ . Let  $\mathcal{H}_t = \{(\pi_0, c_0), (\pi_1, c_0), \dots, (\pi_{t-1}, c_{t-1})\}$  denote the history of recommended items and user's choices on the platform up to time  $t$ . A policy is said to be an online (non-anticipating) policy if the recommendation at time  $t$  depends only on the history  $\mathcal{H}_t$  up

till time  $t$ . We denote the set of all online policies as  $\Pi$ . For any online policy  $\pi \in \Pi$ , we denote  $\text{Eng}(\pi)$  and  $\text{Util}(\pi)$  as the expected engagement and expected utility achieved under the policy  $\pi$ .

$$\text{Eng}(\pi) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \delta^t \mathbb{1}\{c_t(\pi_t) \neq \emptyset\} \right] = \sum_{t=0}^{\infty} \delta^t \mathbb{P}(c_t(\pi_t) \neq \emptyset) \quad (4.3)$$

$$\text{Util}(\pi) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \delta^t \max_{j \in \pi_t \cup \{\emptyset\}} u_j \right] = \sum_{t=0}^{\infty} \delta^t \mathbb{E} \left[ \max_{j \in \pi_t \cup \{\emptyset\}} u_j \right] \quad (4.4)$$

Note that the expectation is also taken over the distribution of the base utility of the niche type.

### 4.3 Analysis of the Two-point distribution for the niche type

In this section, we will focus on a two point distribution for base utility of the niche type and assume that this two-point distribution  $F_N$  is known to the platform. In particular, we will assume that

$$\mathbb{P}(V_N = (1 - p)/p) = p, \quad \mathbb{P}(V_N = -1) = 1 - p \quad (4.5)$$

This implies that the mean base utility of the niche type is zero, with a fraction  $p$  of users valuing it highly at  $V_N = (1 - p)/p$  and the rest valuing it at  $V_N = -1$ . Furthermore, we will assume that the base utility of the popular product  $V_P$  is positive and known to the platform. To start, we study such a setting where the platform has some knowledge about the base utility of the item types, to highlight the utility loss that occurs when platforms optimize for imperfect proxies of utility such as engagement. We will relax these assumptions in Section 4.4 where we will not assume knowledge of the base utility of the popular or niche types and we will observe that our insights continue to hold in many parametric regimes.

**APP: Engagement Optimal Policy** Given that on average the popular type generates higher utility than the niche type, a reasonable baseline policy to consider is a “greedy” policy, which recommends both items of the popular type, dubbed Always Popular Policy, APP in short. This

policy is intimately related to the ranking-based algorithms typically deployed in practice, where items are sorted based on their predicted engagement. Since *on average across the population*, the popular type generates higher engagement than the niche type due to higher mean base utility, applying the ranking-based logic would result in recommending items only of the popular type. We formally define APP as

$$\text{APP} = ((\pi_t^{\text{APP}})_{t \geq 0}), \quad \pi_t^{\text{APP}} = \{i_{1,t}, i_{2,t}\} \text{ s.t. } \tau(i_{j,t}) = \text{P}, \forall j \in \{1, 2\}. \quad (4.6)$$

Note that since APP never recommends an item of the niche type, it is unable to discern whether or not the niche type is preferred over the popular type by a particular user. This presents us with the classical exploration-exploitation dilemma, where in order to learn whether the niche type generates higher utility (and engagement) than the popular type, the platform necessarily needs to explore and recommend items of the niche type, however doing so could potentially hurt engagement. It turns out if the fraction of users  $p$  who derive high utility from the niche type is small enough, then APP is indeed engagement maximizing. This is formalized below.

**Theorem 7 (Engagement Optimal Policy)** *Fix the base utility of the popular item type  $V_P \in \mathbb{R}_+$  and the discount factor  $\delta \in [0, 1)$ . There exists a  $p_0 = p_0(\delta, V_P) \in [0, 1]$  such that for all  $p \leq p_0$ , then APP as defined in (4.6) maximizes engagement as defined in (4.3). Moreover, the expected engagement and utility under APP is*

$$\begin{aligned} \text{Eng}(\text{APP}) &= \frac{1}{1 - \delta} \cdot \frac{2e^{V_P}}{1 + 2e^{V_P}}, \\ \text{Util}(\text{APP}) &= \frac{1}{1 - \delta} \cdot \ln(1 + 2e^{V_P}). \end{aligned}$$

The proof of Theorem 7 is deferred to Appendix C.2. Next, we will discuss an exploration-based utility-aware heuristic, and characterize in closed form its expected engagement and utility.

**PEAR: A utility-aware heuristic** While exploration hurts engagement, it can lead to substantial gains in terms of utility. To this end, we design a simple utility-aware heuristic called Posterior-

based Exploration-driver Adaptive Recommendations, PEAR in short, which initially recommends a diverse set of recommendations consisting of one item of the popular type and the other item of the niche type. The utility-aware heuristic PEAR maintains a posterior belief on the probability of the niche utility being  $(1 - p)/p$  and switches to recommending both items of the popular type when (and if) this posterior belief falls below the initial prior belief  $p$ . The exploration enables the recommendation system to learn whether the niche type generates high utility for the user or not. We formally describe PEAR in Algorithm 4.

---

**Algorithm 4:** Posterior-based Exploration-driven Adaptive Recommendations (PEAR)

---

**Input:** Base utility of the popular item  $V_P$ , Parameter  $p$  in the niche type distribution (defined in (4.5))

**Initialize:**  $p_0 \leftarrow p$ ,  $S \leftarrow 0$ ,  $F \leftarrow 0$ ,  $\rho_1 \leftarrow \frac{e^{(1-p)/p}}{1+e^{V_P+e^{(1-p)/p}}}$ ,  $\rho_2 \leftarrow \frac{e^{-1}}{1+e^{V_P+e^{-1}}}$ .

**for**  $t \in \mathbb{N}$  **do**

$\triangleright$  recommend diverse items till posterior at least  $p$

**if**  $p_t \geq p$  **then**

$\pi_t = \{(i_1, i_2) : \tau(i_1) = \mathbf{P} \text{ and } \tau(i_2) = \mathbf{N}\}$   $\triangleright$  diverse recos

**if**  $\tau(c(\pi_t)) = \mathbf{N}$  **then**

$S \leftarrow S + 1$   $\triangleright$  increment success counter

**else**

$F \leftarrow F + 1$   $\triangleright$  increment failure counter

**end**

**else**

$\pi_t = \{(i_1, i_2) : \tau(i_1) = \mathbf{P} \text{ and } \tau(i_2) = \mathbf{P}\}$   $\triangleright$  only popular recos

**end**

$p_{t+1} = \left(1 + \frac{1-p}{p} \cdot \frac{\rho_2^S (1-\rho_2)^F}{\rho_1^S (1-\rho_1)^F}\right)^{-1}$   $\triangleright$  update posterior

**end**

---

**Theorem 8 (Analysis of PEAR)** Fix the base utility of the popular item type  $V_P \in \mathbb{R}_+$  and discount factor  $\delta \in [0, 1)$ . Define  $\rho \triangleq 1/(1 + e + e^{V_P+1})$ . As  $p \rightarrow 0$ , the expected engagement and utility under PEAR (Algorithm 4) is given as

$$\text{Eng(PEAR)} = \frac{1}{1 - \delta\rho} \cdot \frac{e^{V_P} + e^{-1}}{1 + e^{V_P} + e^{-1}} + \frac{1}{1 - \delta} \cdot \frac{\delta(1 - \rho)}{1 - \delta\rho} \cdot \frac{2e^{V_P}}{1 + 2e^{V_P}},$$

$$\text{Util(PEAR)} = \frac{1}{1 - \delta} + \frac{1}{1 - \delta\rho} \cdot \ln(1 + e^{-1} + e^{V_P}) + \frac{1}{1 - \delta} \cdot \frac{\delta(1 - \rho)}{1 - \delta\rho} \cdot \ln(1 + 2e^{V_P}).$$

The proof of Theorem 8 is deferred to Appendix C.3. Characterizing the utility maximizing policy is technically challenging however we show in Corollary 7 that PEAR is approximately utility optimal for a very forward looking platform ( $\delta \rightarrow 1$ ).

**Corollary 7 (Asymptotic Optimality of PEAR)** *Let  $\text{Util}(\text{OPT})$  denote the optimal utility obtained by an oracle who at time  $t = 0$  knows whether a user prefers niche over popular items or not (and recommends only the preferred item type). Then,*

$$\lim_{\delta \rightarrow 1} \lim_{p \rightarrow 0} \frac{\text{Util}(\text{PEAR})}{\text{Util}(\text{OPT})} = 1.$$

Using Theorems 7 and 8, it follows that for any discount factor  $\delta \in [0, 1)$  and  $V_P > -1$ , in the limit  $p \rightarrow 0$ , we have that,

$$\text{Eng}(\text{PEAR}) < \text{Eng}(\text{APP}), \quad \text{Util}(\text{PEAR}) > \text{Util}(\text{APP}).$$

The above set of inequalities illuminate the misalignment between engagement and utility. Next, we discuss the magnitude of this misalignment, especially in the case when the platforms become increasingly forward looking.

**Asymmetry in the engagement-utility misalignment** As the platforms become increasingly forward looking, i.e.,  $\delta \rightarrow 1$ , we show that engagement under the utility-aware heuristic PEAR approaches the optimal engagement achieved by APP. In contrast we show that there is a non-vanishing gap between the utility obtained by the utility-aware heuristic PEAR and the engagement maximizing policy APP. This result is formalized in Corollary 8 below.

**Corollary 8 (Asymmetry in the Misalignment)** *Fix the attraction parameter of the popular item*

type  $V_P \in \mathbb{R}_+$ . Consider the parameter  $p$  in (4.5) and the discount factor  $\delta$ . Then we have that,

$$\lim_{\delta \rightarrow 1} \lim_{p \rightarrow 0} \frac{\text{Eng}(\text{PEAR})}{\text{Eng}(\text{APP})} = 1 \quad (4.7)$$

$$\lim_{\delta \rightarrow 1} \lim_{p \rightarrow 0} \frac{\text{Util}(\text{PEAR})}{\text{Util}(\text{APP})} = 1 + \frac{1}{\ln(1 + 2e^{V_P})} \quad (4.8)$$

Corollary 8 follows from Theorems 7 and 8. As  $\delta \rightarrow 1$ , from (4.8), we observe that there is substantial utility gain by the utility-aware heuristic PEAR with respect to the engagement maximizing policy APP. From (4.8), it follows that

$$\lim_{\delta \rightarrow 1} \lim_{p \rightarrow 0} \frac{\text{Util}(\text{PEAR}) - \text{Util}(\text{APP})}{\text{Util}(\text{APP})} \geq \frac{1}{V_P + 1}, \quad \forall V_P \geq \frac{1}{3}$$

For  $V_P = 1$ , the above inequality shows that the relative gain in utility of PEAR over APP is approximately 50% which is line with the observations in Figure 4.1 and Table 4.1 presented in the introduction. While characterizing the expected utility and engagement of the utility optimal policy is technical challenging, we expect the utility optimal policy to have similar performance to that of PEAR – small loss in engagement and large gain in utility compared to APP.

Finally, we acknowledge that in this section, we make certain simplifying assumptions such as knowledge of the base utility distribution of the two types, the two-point distribution of the base utility of the niche type and focus on the limiting regime of  $p \rightarrow 0$ . These simplifying assumptions allow us to distill the key structural insights via closed form characterizations, with minimal dependence on different parameters of the model. In the following section, we will relax these assumptions and numerically demonstrate that many of our insights continue to hold under varied settings.

#### 4.4 Robustness of Insights Under General Settings

In this section, we will relax the assumptions of Section 4.3. In particular, we relax the assumption that the mean utility of the popular type and the utility distribution of the niche type is

known to the platform. We will restrict our attention to two policies: (i) APP as defined in (4.6), and (ii) the Diverse-then-Customized (DICE in short) heuristic with an exploration parameter  $\mathcal{T}$ . DICE is an explore-then-commit type policy studied in the multi-armed bandit literature [121]. The DICE heuristic is a *prior-free* heuristic which recommends a mix of popular and niche items in the first  $\mathcal{T}$  periods and then switches to recommends a customized homogeneous set of recommendation (either popular or niche type) based on the user choices in the first  $\mathcal{T}$  periods. We formally describe the DICE heuristic in Algorithm 5. We will numerically analyze the case when the base utility of the niche type is drawn from a generalized Pareto distribution with parameters  $\mu, \sigma$  and  $\xi$  and denoted as  $\text{GPD}(\mu, \sigma, \xi)$ . We will focus on a special class of these distributions where we fix  $\mu = -1$  and  $\sigma = 1 - \xi$  and we only have a single parameter  $\xi$ ; we will denote these distributions as  $F_N^\xi$ . For any  $\xi < 1$ , the mean of the distributions is zero (same as the outside option). The CDF of the distributions as a function of the parameter  $\xi$  is provided below.

$$F_N^\xi(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{1-\xi} \cdot (x+1)\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-(x+1)), & \xi = 0 \end{cases}$$

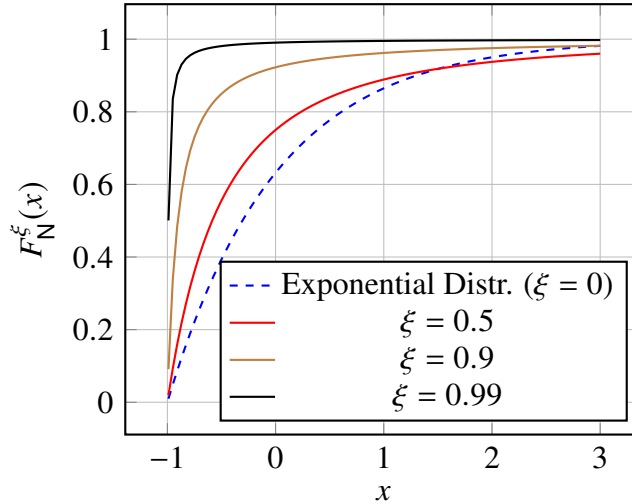


Figure 4.2: CDF of Generalized Pareto Distribution for  $\mu = -1$  and  $\sigma = 1 - \xi$

For  $\xi = 0$ , we have the exponential distribution and for  $\xi > 0$ , we have the Pareto distribution

---

**Algorithm 5: Diverse-then-CustomizEd (DICE)**

---

**Input:** Exploration Phase Length  $\mathcal{T}$   
**Initialize:**  $\text{Count}_P \leftarrow 0, \text{Count}_N \leftarrow 0$   
**for**  $t \in \{1, 2, \dots, \mathcal{T}\}$  **do**  
     $\pi_t = \{(i_1, i_2) : \tau(i_1) = P \text{ and } \tau(i_2) = N\}$   $\triangleright$  diverse recos  
    **if**  $\tau(c(\pi_t)) = P$  **then**  
         $\text{Count}_P \leftarrow \text{Count}_P + 1$   $\triangleright$  increment popular counter  
    **else if**  $\tau(c(\pi_t)) = N$  **then**  
         $\text{Count}_N \leftarrow \text{Count}_N + 1$   $\triangleright$  increment niche counter  
    **else**  
        continue  
    **end**  
**end**  
 $\triangleright$  check if user chooses more of popular or niche type  
**if**  $\text{Count}_P \geq \text{Count}_N$  **then**  
     $\text{PrefType} = P$   
**else**  
     $\text{PrefType} = N$   
**end**  
 $\triangleright$  recommend the more chosen type during exploratory phase  
     $\{1, 2, \dots, \mathcal{T}\}$   
**for**  $t \in \{\mathcal{T} + 1, \mathcal{T} + 2, \dots\}$  **do**  
     $\pi_t = \{(i_1, i_2) : \tau(i_1) = \tau(i_2) = \text{PrefType}\}$   $\triangleright$  customized recos based on users  
        choices in exploratory phase  
**end**

---

with scale parameter  $\xi$ , which is a heavy-tailed distribution. From Figure 4.2, we observe that as  $\xi$  increases, the probability of the niche type having positive base utility decreases. Note that since we fix the mean of the distribution to be zero, increasing  $\xi$  corresponds to the tail becoming lighter, i.e.,  $\bar{F}_N(V) = 1 - F_N(V)$  is decreasing as  $\xi$  increases for a fixed  $V \geq -1$ . Increasing  $\xi$  is analogous to the case of decreasing  $p$  to zero for the two-point distribution of the niche type as defined in (4.5). We compare the expected engagement and utility of APP and DICE in Figures 4.3 and 4.4 for  $\delta = 0$  and  $\delta = 0.999$  respectively. As the baseline, we consider APP and the bar plots in Figures 4.3 and 4.4 demonstrate the relative change in engagement (in blue) and utility (in red) under DICE compared to APP. If a given metric (engagement or utility) is below/above the baseline APP, it implies that DICE loses/gains compared to the APP with the percentage loss or gain depicted by the height of the bar plot. The key insights are as follows:

1. For a fixed discount factor  $\delta \in [0, 1)$ , for larger tail parameter  $\xi$  values (closer to 1), we observe misalignment between engagement and utility optimizing policies. In particular, we observe in Figures 4.3 and 4.4, that  $\text{Eng}(\text{APP}) > \text{Eng}(\text{DICE})$  and  $\text{Util}(\text{APP}) < \text{Util}(\text{DICE})$  for large enough  $\xi$ .
2. As  $\delta$  approaches 1, i.e., as the platforms become more forward looking, we observe a stark asymmetry in the magnitude of the misalignment. In Figure 4.4, we observe the there is little difference in the engagement obtained by DICE and APP. However, there is a significant enhancement in utility under DICE compared to APP.

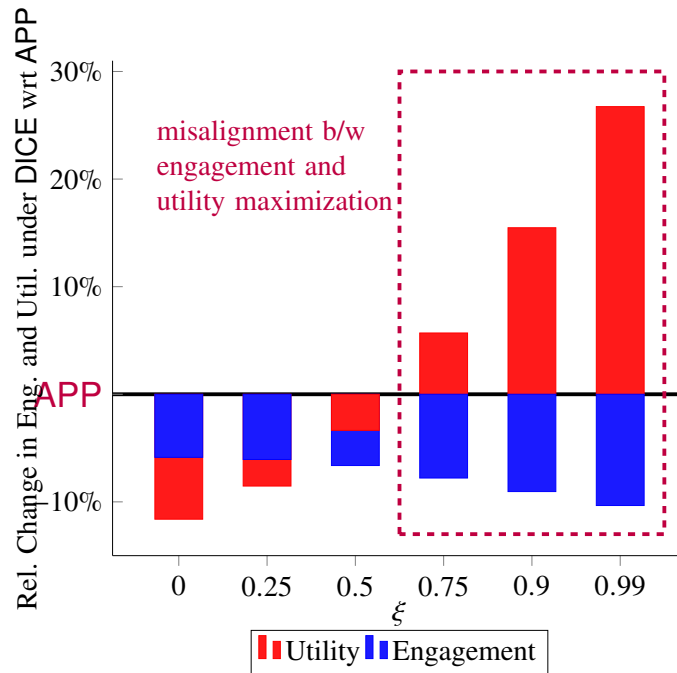


Figure 4.3: Impact of  $\xi$  on utility and engagement for  $\delta = 0$

## 4.5 Conclusion

In this work, we explore the extent of user utility loss when platforms use measurable but imperfect proxies like engagement to drive recommendations, and whether it is feasible to optimize for user utility, which is rarely measured. We studied a model where a recommendation system repeatedly suggests items to a user, adapting its selections over time. Our findings reveal a

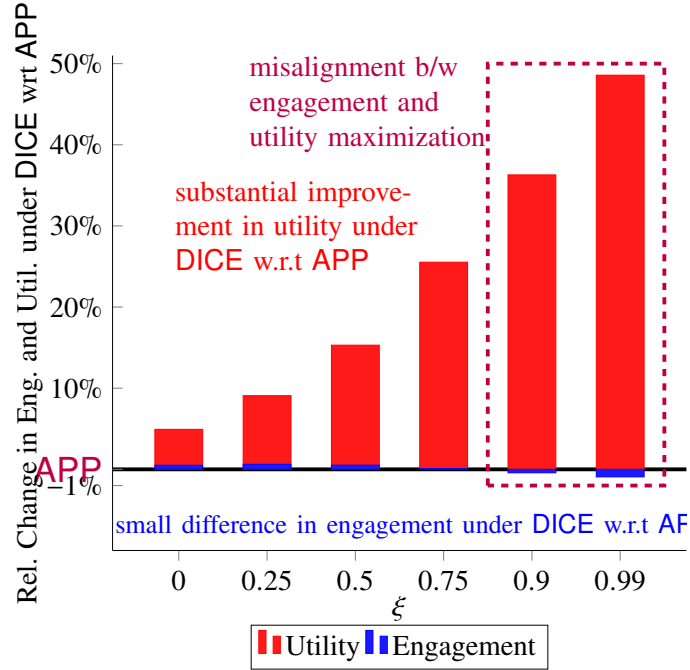


Figure 4.4: Impact of  $\xi$  on utility and engagement for  $\delta = 0.999$

fundamental misalignment between policies that maximize engagement and those that maximize utility. We develop and study a utility-aware heuristic PEAR which is able to achieve best of both worlds: near-optimal user utility and near-optimal engagement simultaneously. This highlights a stark asymmetry in the misalignment: substantial improvement in utility is achievable in comparison to the engagement-optimal policy APP while sacrificing a minuscule amount on engagement. Moreover, we observe that some of these insights also carry over to the setting where the platform has no prior information about the utility distribution. Overall, our research highlights that recommendation systems with the ability to recommend more than one item can facilitate exploration with minimum reduction in engagement, allowing discovery of items with higher utility and hence leading to a significant enhancement in user utility.

Our model is intentionally simplified and does not encompass all aspects of contemporary recommendation systems. This simplicity allows us to focus on key tension between engagement and utility, however, our model may be expanded upon to study different facets of the recommendation systems. Amongst many open directions, our model can potentially be extended to (i) incorporate the setting with many niche types, (ii) model user satiation, where the utility at each time depends

not only on the current consumption but past consumption as well, (iii) time-varying preferences and/or inconsistent preferences. We defer these extensions and other exciting open questions for future research.

## Chapter 5: Impact of Rankings and Personalized Recommendations in Marketplaces

*Based on the paper [122] co-authored with Omar Besbes and Yash Kanoria.*

### 5.1 Introduction

Every day, individuals navigate choices ranging from the mundane (e.g., selecting a movie) to life-altering (e.g., choosing a college). Surveys and empirical research have demonstrated, and personal experience affirms, that these decisions are frequently made under *imperfect information*, where preferences are rarely fully formed, and outcomes often lead to regret. A 2017 survey found that most U.S. adults would alter their educational choices if given the chance [123], underscoring a broader phenomenon: individuals often have poorly formed preferences when making choices.

To help guide decisions, *public rankings* have emerged as a ubiquitous tool across many domains. In entertainment, websites like IMDb and Billboard rank movies and songs by popularity, respectively. In e-commerce, in their early days, platforms like Amazon used bestseller lists and star ratings to highlight popular products. In education, organizations such as the US News & World Report [124], and the National Institutional Ranking Framework (NIRF) [125] in India, rank universities and colleges based on various performance metrics. These rankings aggregate information and simplify decision-making, serving as a common signal of quality in settings where individuals struggle to evaluate options on their own. Rankings are particularly influential in settings where users lack direct experience – for example, prospective college students who have never attended the institutions they are choosing between. However, while rankings provide a useful population-level signal, they fail to capture *idiosyncratic preferences*, i.e., the way in which the individual user’s value for an item differs from the average user’s value for that item.

The rise of *personalized recommendation tools* in several domains has provided an alternative approach, tailoring choices to individual preferences rather than presenting a single, universal ranking. In entertainment, platforms like Netflix, Spotify, and YouTube curate recommendations based on user behavior, revealing content that aligns with individual tastes. In e-commerce, platforms like Amazon and Etsy now personalize search results, increasing sales by showing products aligned with past browsing and purchase behavior. Similarly, in higher education, platforms like Naviance and Scoir use historical data and student profiles to recommend universities that align with a student's specific strengths and interests. These systems go beyond general quality signals to help individuals find the best choices for them, rather than just the highest-ranked options. With the rise of generative AI, it is becoming increasingly common to see chatbots that personalize recommendations. In e-commerce, Amazon's Rufus, an AI-powered shopping assistant, engages in real-time conversations with users, answering open-ended queries and generating personalized product suggestions based on Amazon's extensive catalog and user behavior. Similarly, in college admissions, platforms like CollegeVine and Kollegio are leveraging AI to provide personalized counseling to students, offering tailored recommendations for universities. These tools bring a new dimension to recommendation systems by enabling dynamic, dialogue-based interactions rather than static ranked lists.

Despite the prevalence of personalized recommendation tools in e-commerce and entertainment, there are still many domains where this technology has not achieved widespread penetration. Policymakers and platform designers, for example, may ask how much value these tools truly deliver and in which settings they add the most value relative to public rankings. A concrete case study arises in college admissions: the Indian government invests substantially in NIRF to evaluate universities. Would it be more or less beneficial to invest in developing and deploying personalized recommendation services that help match students to programs that reflect their individual preferences and needs? These questions extend beyond college admissions. The design of Netflix's recommendation system, Airbnb's ranking algorithms, or similar platforms also involves balancing broad quality signals against more tailored, individual-specific information. As investment in

AI-powered personalized recommendation systems grows – particularly in light of advances in generative AI [126, 127] – it becomes vital to understand when and how these tools deliver more (or less) value than traditional public rankings.

These considerations lead to the following key policy and design questions: Should a designer seeking to improve welfare emphasize high-quality public rankings, or invest in personalized recommender systems? How does the answer differ when supply constraints exist (e.g., limited seats in college programs or a given inventory of listings on a lodging marketplace) versus environments where supply is effectively unconstrained? Given that in many domains, advanced AI-driven personalization tools will soon be feasible to build, one may ask how much additional societal value would such advanced tools provide over and above that provided by existing public ranking tools? Succinctly put, we ask the following question in this paper:

*What are the implications of different information provisioning tools, such as public rankings and personalized recommendations, in environments with and without supply-side constraints?*

It is unsurprising that personalized recommendations outperform public rankings; that is not what we aim to study. Instead, we aim to quantify both the incremental value that personalization offers over public rankings alone, and the benefit public rankings provide relative to having no information provisioning tool at all. In doing so, we identify the core drivers of these gains and examine how they play out in different environments. To isolate these effects cleanly, we study a stylized, parsimonious model that captures the essential features of the information-provisioning tools and the different market settings, while abstracting away from the specifics of the operationalization of these tools. We outline the model’s basic ingredients here, with the formal description deferred to Section 5.2.

**Role of information provisioning tools** The agents’ utility for an item is modeled as a weighted combination of two *independently drawn* terms: a *common term* which depends only on the item,

reflecting the item’s population-level quality and an *idiosyncratic term* which depends on the agent-item pair and captures agent-specific adjustments. We weight the common and idiosyncratic terms using the parameters  $1 - \rho$  and  $\rho$ , respectively, where  $\rho \in [0, 1]$  reflects the level of heterogeneity. Lower values of  $\rho$  indicate that the utility of agents is mainly driven by the common term shared between agents, while higher values of  $\rho$  imply greater influence of the idiosyncratic term, reflecting more heterogeneous preferences. We model the role of the different information provisioning tools as informing agents of different components of their utility: public rankings inform only the common term, while personalized recommendations reveal both the terms.

**Uncapacitated and capacitated supply settings** We assume that there are  $n$  agents and  $n$  items.

We assume that each agent has unit demand, i.e., consumes only a single item.

- (i) *Uncapacitated Supply Setting*. This setting is motivated by content recommendation platforms like Netflix, Spotify and Youtube, where there is no restriction on the number of agents who can consume a given item. To capture this, we assume that each item has infinite capacity.
- (ii) *Capacitated Supply Setting*. This setting is motivated by online marketplaces like Airbnb as well as centralized college admissions, where supply is constrained. Agents choose amongst a multitude of items, and we model capacity constraints by assuming that each item can be matched to at most one agent.

To isolate and quantify the marginal impact of these two information provisioning tools, we study three different information regimes (see Figure 5.1): (i) **No Information** (denoted  $\emptyset$ ) where agents lack knowledge of both the common as well as the idiosyncratic terms, (ii) **Only Quality Information** (denotes  $q$ ) where public rankings provide agents with the common terms only, and (iii) **Full Information** (denoted  $u$ ) where agents have access to both the common and the idiosyncratic term through personalized recommendations.

We measure goodness of outcomes in terms of social welfare of agents, which we quantify by the average utility across agents, termed average welfare (**AW**), obtained under different informa-

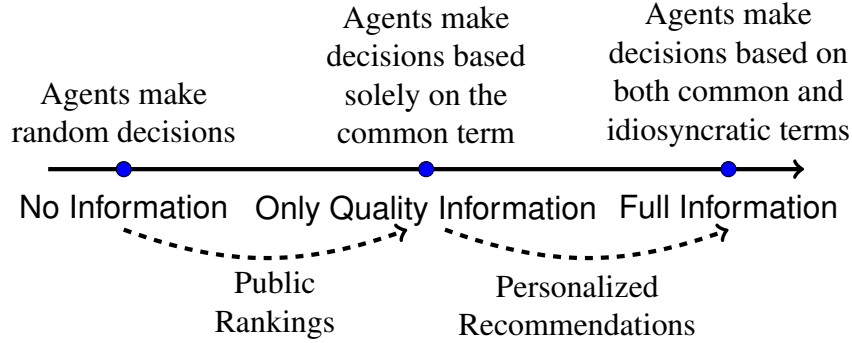


Figure 5.1: Different information regimes studied in this work

tion regimes and different environments. We assume that the common terms and the idiosyncratic terms are drawn independently from distributions  $P_q$  and  $P_\varphi$  respectively. Motivated by empirical findings, we primarily focus on distributions with Pareto tails, reflecting the prevalence of power-law behavior in measures of popularity and success [128]. As a special case of Pareto tails, we also consider distributions with exponential tails [101].

### 5.1.1 Main Contributions

In this work, we develop a stylized and parsimonious model to examine the interplay between information provisioning tools and different market environments. Through this model, we isolate key value drivers, offering insights for practitioners and policymakers. On the technical front, we characterize the value of public rankings and personalized recommendations in large markets with Pareto and exponential-tailed distributions. Our key contributions are in formulating a parsimonious model and the crisp insights that follow as a result. We now elaborate on our contributions.

- *Fundamental Role of Capacity and Heterogeneity.* We identify that both (i) capacity constraints and (ii) level of preference heterogeneity (captured by the parameter  $\rho$ , the weight of the idiosyncratic utility term) play a key role in determining the value of different information provisioning tools. In Figure 5.2, we illustrate the different asymptotic “rates” or scaling of welfare gain across these regimes, highlighting the interplay of level of heterogeneity and supply constraints on the marginal impact of each of these tool. Although we introduce the

asymptotic rates here for an at-a-glance overview, its details and proofs are developed fully in Section 5.3.

- *Uncapacitated Setting.* In the absence of capacity constraints, both public rankings and personalized recommendations improve aggregate agent utility, with their relative value hinging critically on  $\rho$  (level of heterogeneity). If  $\rho$  is small, i.e., agent preferences align closely with the common utility term, public rankings capture the bulk of the welfare gains, as they reveal this shared component (Figure 5.2, first row, left column). Conversely, if  $\rho$  is large, i.e., preferences are mostly driven by the idiosyncratic utility term, personalized recommendations add greater value by additionally revealing the idiosyncratic term, thereby tailoring information to individual agents (Figure 5.2, first row, right column).
- *Capacitated Setting.* In stark contrast, in the capacity constrained setting, revealing just the common term through public rankings provides no value in aggregate. Personalized recommendations do generate value, by accounting for the idiosyncratic term of the utility. As before,  $\rho$  drives the value generated by personalized recommendations – a larger value of  $\rho$  correspond to larger welfare improvement by personalizing recommendations (Figure 5.2, second row).

The distinction arises from the dual role of personalized recommendations in these settings. Public rankings identify the best overall options, providing agents with population-level insights into item quality. Personalized recommendations, however, go further: they (i) refine agents’ preferences by revealing individualized utility components and (ii) improve the allocation of agents to items. In capacity-constrained settings, such as online marketplaces or college admissions, both of these effects are crucial. Conversely, in unconstrained environments, such as content recommendation platforms, the primary benefit of personalized recommendations lies in preference refinement, as allocation considerations are irrelevant. This dichotomy highlights a fundamental interplay between supply-side constraints and the

value of information provisioning tools.

- *Characterization of welfare gains.* We formally derive how welfare scales with market size  $n$  under Pareto and exponential-tailed distributions (Theorems 9, 10, 11, and 12).

For the uncapacitated setting, the key technical challenge lies in characterizing the additional welfare gain due to personalizing recommendations. This requires characterizing the tail behavior of random variable which is a weighted combination of two random variables with Pareto and exponential tails. While the analysis is not too involved, our result highlights interesting asymmetric impact of the information provisioning tools: for  $\rho \in (0, 1/2)$ , public rankings (revealing the common utility term) account for most welfare gains from recommendations, with minimal benefits from upgrading to personalized recommendations. For  $\rho \in (1/2, 1)$ , personalizing recommendations (revealing the idiosyncratic term) contribute most value (Figures 5.3b, 5.3c, 5.3d). This asymmetry is most pronounced for exponential-tailed distributions, where a phase transition occurs (Figure 5.4b): personalizing recommendations yield no additional value for  $\rho \in (0, 1/2)$  but drives significant gains for  $\rho \in (1/2, 1)$ .

For the capacitated setting, the main technical challenge lies in characterizing the welfare gains from personalized recommendations. We circumvent this key challenge by providing a lower and upper bound on the welfare gains in Lemma 4 and show that these bounds are asymptotically tight for the Pareto and exponential tailed distributions. However, for the case of bounded distributions, closing the gap between the upper and lower bounds is challenging (see Appendix D.3) and we defer this question for future research. Our analysis shows that the additional welfare gains due to personalized recommendations scale with the level of heterogeneity: large  $\rho$  corresponds to larger benefits of personalizing recommendations (see Figure 5.2, second row).

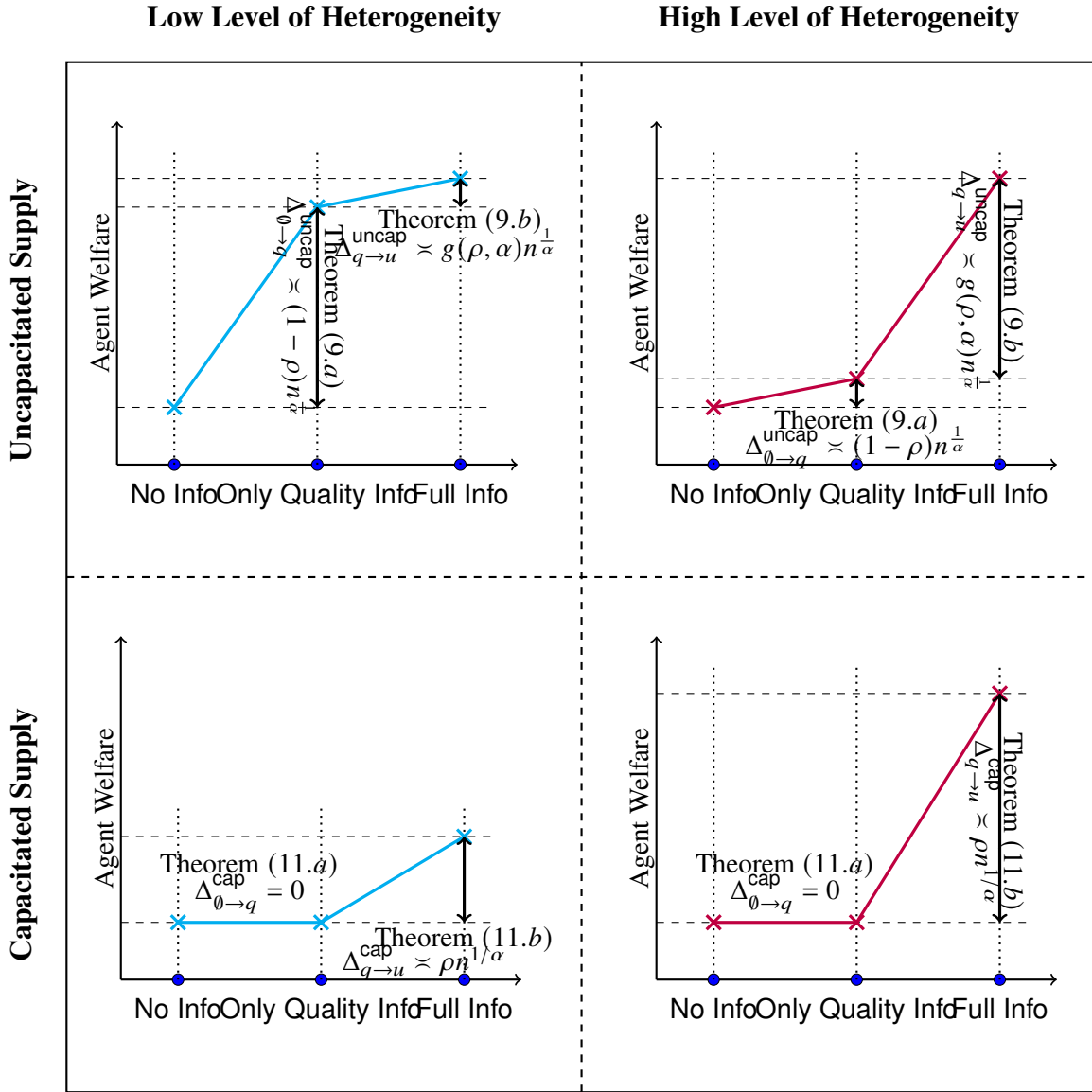


Figure 5.2: Shows the marginal impact of public rankings and personalized recommendations and their interplay with (i) capacity constraints (in the rows) and (ii) level of heterogeneity (in the columns). Low level of heterogeneity refers  $\rho \in (0, 1/2)$  and high level of heterogeneity refers  $\rho \in (1/2, 1)$ .

### 5.1.2 Related Literature

This work is motivated by and contributes to several strands of literature on recommendation systems, matching with incomplete preferences, and information design in matching markets.

*Recommendation Systems and Decision Support Tools.* Classical recommendation systems have focused on identifying and suggesting items that best fit each user’s preferences [129, 56]. These recommendation and decision support tools have shown great promise in terms of improving the decisions made by users [130]. The emphasis has been on developing accurate user behavior model and develop methods to improve the relevance of personalized recommendations [131, 132, 133]. These methods have mostly been designed to operate in uncapacitated environments, such as content streaming platforms, and as such do not generally take into account matching or capacity constraints. More recently, motivated by e-commerce and labor platforms, there has been a growing interest in designing recommendation systems which take into these matching constraints [134, 135, 136]. The focus of these papers has been methodological while in this work, we aim to understand the nuanced interplay between supply side capacity constraints and the value that personalized recommendations can generate.

*Incomplete Preferences and Informational Interventions in Matching Markets.* Most of the literature on one-sided and two-sided matching typically assumes that agents possess well-defined preferences [137, 138, 139]. However, these assumptions are often unrealistic in practical scenarios, as recent empirical studies have shown that the absence of well-formed preferences can lead to inefficient matching outcomes [140, 141]. Motivated primarily by applications in school and college admissions, recent work has shifted focus to issues of preference discovery and incomplete information [142, 143, 144]. This body of research typically examines situations where agents strategically acquire additional information to refine their preferences and make informed choices. Empirical and field studies have evaluated the impact of providing additional information to students in the context of high school admissions [145, 146] and college admissions [147, 148]. In particular, [145] provides non-personalized interventions (list of nearby schools with high graduation rates) to students and finds that “*informational interventions may not reduce inequal-*

ity, since both disadvantaged and comparatively advantaged students used our materials”. This finding speaks directly to our insight that in capacitated settings, impersonal tools such as public rankings may not add value in *aggregate*. Our contribution to this line of research takes a modeling approach, aiming to isolate the impact of different information provisioning tools on the average user welfare.

*Information Design in Matching Markets.* There is an emerging literature on information design and signaling in matching markets. This literature typically studies a central platform which chooses to strategically provide information to agents in order to influence their behavior and the resulting matches [149, 150, 151]. In terms of setting, the most closely related paper is [152]. They study the problem where a central planner strategically provisions information to agents with incomplete information in order to optimize for social welfare. A key distinction of this line of work to our work is that we do not study strategic information provisioning rather focus on the impact of different information provision tools.

*Algorithmic monoculture and homogenization.* [153] first formalized algorithmic monoculture in hiring markets, where many firms assess applicants with the *same* ranking algorithm. By contrast, algorithmic polyculture describes settings in which firms rely on independent algorithms. Extending these ideas to large two-sided matching markets, [154] analyze stable matching outcomes under monoculture and polyculture and show that, when evaluation noise is well behaved, monoculture can reduce firm utility by resulting in less preferred applicants being hired vis-a-vis polyculture. [155] studies the impact of noise in the evaluation of candidates to the resulting stable matching outcome in the context of polyculture – in particular, they consider the impact of the tail of the noise distribution on the outcome. Our capacitated model maps directly onto these notions. In the Only Quality Information regime, all agents share a single impersonal ranking—mirroring monoculture, whereas in the Full Information regime each agent has an individualized ranking, paralleling polyculture. Similar to [154], we find that the agent welfare is lower in the Only Quality Information regime (monoculture) compared to the Full Information regime (polyculture). A recent work by [156] incorporates strategic behavior into the monoculture setting and character-

ize the resulting Nash equilibria. While our work does not study strategic behavior on part of the agents, unlike [156], qualitatively speaking, our insights resonate with [156]: (i) competition for the top items (candidates in [156]) leads to inefficiencies due to congestion or matching constraints and (ii) in the capacitated setting, most of the value lies in matching agents (firms) to items (candidates) that they idiosyncratically value highly.

## 5.2 Model

We consider a balanced market with  $n$  agents (set  $\mathcal{X}$ ) and  $n$  items (set  $\mathcal{Y}$ ). Each agent  $x \in \mathcal{X}$  has a unique priority score  $s_x \in \mathbb{R}$ . The utility of agent  $x$  for item  $y$  is given as

$$u_{xy} = (1 - \rho) q_y + \rho \varphi_{xy}, \quad \forall x \in \mathcal{X}, y \in \mathcal{Y} \quad (5.1)$$

where,  $q_y$  and  $\varphi_{xy}$  are independent terms and,

- $q_y$  is a common term which depends only on item  $y$ , drawn i.i.d from a distribution  $P_q$ .
- $\varphi_{xy}$  is an idiosyncratic term for the agent-item  $(x, y)$  pair, drawn i.i.d from a distribution  $P_\varphi$ .
- $\rho \in [0, 1]$  is a parameter that determines the relative weight of the idiosyncratic term, capturing level of heterogeneity. Smaller  $\rho$  implies more homogeneous preferences (common term dominates), while larger  $\rho$  implies more heterogeneous preferences (idiosyncratic term dominates).

Agents select items sequentially in  $n$  rounds, ordered by their priority scores (highest score chooses first, etc.). In round  $k$ , the  $k$ -th agent chooses from the remaining items (denoted as  $\mathcal{Y}_k^{\text{rem}}$ ) to maximize her perceived utility, with ties broken uniformly at random<sup>1</sup>. We study three

---

<sup>1</sup>This model encompasses the main examples of interest. In the uncapacitated setting, the sequence does not matter because items have infinite capacity. In the capacitated case, a priority-based order aligns with centralized college admissions, where students are ranked by an exam score and sequentially pick from available programs [157, 137, 158]: in the balanced market setting with common preferences on the supply side, deferred acceptance is equivalent to serial dictatorship. If priority scores are random, this corresponds to the random arrival model in online marketplaces.

information regimes as mentioned below. Let  $\sigma_\star(k)$  be the index of the item chosen by the  $k$ -th agent in regime  $\star \in \{\emptyset, q, u\}$ .

- (i) **No Information ( $\emptyset$ ):** The agent has no information about any items, perceives all items as identical, and hence chooses uniformly at random among the remaining items.
- (ii) **Only Quality Information ( $q$ ):** The agent only knows the common term ( $q_y$ ) and agent  $k$  chooses the item with the highest value of the common term, since the idiosyncratic term for all the items is the same from the agent's point of view. In particular, we have that

$$\sigma_q(k) \triangleq \arg \max_{y \in \mathcal{Y}_k^{\text{rem}}} (1 - \rho)q_y + \rho\varphi_{ky} = \arg \max_{y \in \mathcal{Y}_k^{\text{rem}}} q_y,$$

where  $\varphi_{ky} = 0, \forall y \in \mathcal{Y}_k^{\text{rem}}$  since the agents have no information about the idiosyncratic term.

- (iii) **Full Information ( $u$ ):** The agent knows both the common terms ( $q_y$ ) as well as the idiosyncratic terms ( $\varphi_{xy}$ ) and agent  $k$  chooses the item with the highest utility. In particular, we have that

$$\sigma_u(k) \triangleq \arg \max_{y \in \mathcal{Y}_k^{\text{rem}}} (1 - \rho)q_y + \rho\varphi_{ky}.$$

We study two types of supply constraints:

- (a) Uncapacitated Supply: Each items has *infinite capacity*; any number of agents can choose the same item.
- (b) Capacitated Supply: Each item has *unit capacity*; once chosen, it becomes unavailable to subsequent agents.

We define *agent welfare* as the (expected) average utility of agents under each regime.

- Agent Welfare in uncapacitated setting:  $\text{AW}_\star^{\text{uncap}}(n) \triangleq \mathbb{E}[u_{1, \sigma_\star(1)}]$ ,  $\star \in \{\emptyset, q, u\}$ , since all agents effectively face an identical choice as item capacity is infinite.

- Agent Welfare in capacitated setting:  $\text{AW}_\star^{\text{cap}}(n) \triangleq n^{-1} \mathbb{E} \left[ \sum_{k=1}^n u_{k, \sigma_\star(k)} \right]$ ,  $\star \in \{\emptyset, q, u\}$ .

To assess the *marginal value* of public rankings and personalized recommendations, we compare welfare across regimes. For the uncapacitated setting, we have that,

$$\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n) = \text{AW}_q^{\text{uncap}}(n) - \text{AW}_\emptyset^{\text{uncap}}(n), \quad \Delta_{q \rightarrow u}^{\text{uncap}}(n) = \text{AW}_u^{\text{uncap}}(n) - \text{AW}_q^{\text{uncap}}(n).$$

$\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)$  quantify the marginal impact of public rankings and personalized recommendations in the uncapacitated setting, respectively. Similarly, for the capacitated setting, we have that

$$\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n) = \text{AW}_q^{\text{cap}}(n) - \text{AW}_\emptyset^{\text{cap}}(n), \quad \Delta_{q \rightarrow u}^{\text{cap}}(n) = \text{AW}_u^{\text{cap}}(n) - \text{AW}_q^{\text{cap}}(n).$$

We have that  $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{cap}}(n)$  quantify the marginal impact of public rankings and personalized recommendations in the capacitated setting, respectively.

**Notation.** Let  $X$  be a random variable, then  $X_{(k:n)}$  denotes the  $k$ -th order statistic ( $k$ -smallest value) of  $n$  independent and identically distributed copies of  $X$ . Note that  $X_{(n:n)} = \max\{X_1, \dots, X_n\}$  denotes the highest value amongst  $n$  i.i.d draws of  $X$ . For any  $x \in \mathbb{R}$ , we have that  $(x)_+ \triangleq \max\{x, 0\}$ .

### 5.3 Main Results

In this section, we will assume that the common terms  $(q_y)$  and the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from distributions  $P_q$  and  $P_\varphi$ . In order to illuminate the role of the tail of the distribution  $P_q$  and  $P_\varphi$ , we will describe the distributions only in terms of their tails. In particular, we will focus on the Pareto tail (heavy-tailed distribution) which we formally define in Definition 4 below. We also study the case of exponential tail (defined in Definition 5).

**Definition 4 (Pareto Tail)** Fix  $c > 0$  and  $\alpha > 1$ . Let  $X$  be a random variable with distribution  $F$ . We say that  $F$  has a Pareto tail with parameters  $(c, \alpha)$  if  $\lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{(c/x)^\alpha} = 1$ .

**Definition 5 (Exponential Tail)** Fix  $c > 0$  and  $\lambda > 0$ . Let  $X$  be a random variable with distribution  $F$ . We say that  $F$  has an exponential tail with parameters  $(c, \lambda)$  if  $\lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{c \exp(-\lambda x)} = 1$ .

### 5.3.1 Uncapacitated supply setting

Recall that in the uncapacitated supply setting, we have a single agent with unit demand and  $n$  items with unit capacity. Note that agent welfare is simply the expected utility of the item chosen by the agent under different information regimes.

**Theorem 9 (Uncapacitated Supply, Pareto tails)** Consider the uncapacitated supply setting. Fix  $c_q > 0, \alpha_q > 1, c_\varphi > 0, \alpha_\varphi > 1$ . Assume that the common terms  $(q_y)$  are drawn i.i.d from a distribution  $P_q$  with non-negative support, finite mean  $\mu_q < \infty$  and has a Pareto tail with parameters  $(c_q, \alpha_q)$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from a distribution  $P_\varphi$  with non-negative support, finite mean  $\mu_\varphi < \infty$  and has a Pareto tail with parameters  $(c_\varphi, \alpha_\varphi)$ . For any  $\rho \in [0, 1]$ , we have that,

(9.a) The difference in the agent welfare  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  obtained in the Only Quality Information regime and the No Information regime scales in the number of items  $n$  as

$$\lim_{n \rightarrow \infty} \frac{\Delta_{\emptyset \rightarrow \varphi}^{\text{uncap}}(n)}{c_q \Gamma(1 - 1/\alpha_q) \cdot n^{1/\alpha_q}} = 1 - \rho.$$

(9.b) The difference in the agent welfare  $\Delta_{\emptyset \rightarrow \varphi}^{\text{uncap}}(n)$  obtained in the Full Information regime and Only Quality Information regime depends on the values of tail exponents  $\alpha_q$  and  $\alpha_\varphi$  as follows:

(9.b.i)  $\alpha_q \neq \alpha_\varphi$ . Let  $\underline{\alpha} \triangleq \min\{\alpha_q, \alpha_\varphi\}$  and  $c \triangleq c_q \mathbb{1}\{\alpha_q < \alpha_\varphi\} + c_\varphi \mathbb{1}\{\alpha_q > \alpha_\varphi\}$ . Then we have

that

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c\Gamma(1 - 1/\underline{\alpha}) \cdot n^{1/\underline{\alpha}}} = \rho \cdot \mathbb{1}\{\alpha_q > \alpha_\varphi\}.$$

(9.b.ii)  $\alpha_q = \alpha_\varphi$ . Let us denote  $\alpha_q = \alpha_\varphi = \alpha$ . Then we have that,

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} = ((1 - \rho)^\alpha c_q^\alpha + \rho^\alpha c_\varphi^\alpha)^{1/\alpha} - (1 - \rho)c_q.$$

Furthermore, if  $c_q = c_\varphi = c$ , then we have that

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} = ((1 - \rho)^\alpha + \rho^\alpha)^{1/\alpha} - (1 - \rho).$$

The proof of Theorem 9 is deferred to Section 5.4.2. Theorem 9 captures the marginal welfare gains in the uncapacitated setting with Pareto-tailed common and idiosyncratic terms. It is split into two parts:

- Theorem (9.a) quantifies the improvement from No Information to Only Quality Information, showcasing the value of public rankings.
- Theorem (9.b) measures the additional gains from Only Quality Information to Full Information, revealing when personalized recommendations are most beneficial.

### Discussion of Theorem (9.a): Value of Public Rankings

- **Main Insights:** When items have infinite capacity, revealing the common term ( $q_y$ ), as done by public rankings, can significantly improve welfare if  $\rho < 1$  (see Figure 5.3a). Specifically, Theorem (9.a) shows that  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  grows on the order of  $n^{1/\alpha_q}$ , multiplied by  $(1 - \rho)$  and a constant factor related to the parameters of the Pareto tail. Since supply is unlimited, the agent can freely pick the highest- $q_y$  item without being blocked. Because  $(1 - \rho)$  reflects how much the common term contributes to the agent's utility, a smaller  $\rho$  (i.e., more homogeneous

preferences) yields greater benefits from public rankings.

- **Proof Sketch:** In the No Information regime, the agent's expected utility is simply  $(1 - \rho)\mu_q + \rho\mu_\varphi$ . With Only Quality Information, the agent sees the highest  $q_y$ . Because  $q_y$  follows a Pareto tail, its maximum grows like  $n^{1/\alpha_q}$ . This increase is multiplied by  $(1 - \rho)$ , reflecting the weight of the common term in the total utility.

### Discussion of Theorem (9.b): (Incremental) Value of Personalized Recommendations

- **Main Insights:** Theorem (9.b) measures how much additional welfare is gained by revealing both the common and the idiosyncratic terms, rather than only the common term. In the Full Information regime, the agent see both  $(q_y)$  and  $(\varphi_y)$ . Thus, the agent chooses the maximum of  $n$  i.i.d  $Z_y = (1 - \rho) q_y + \rho \varphi_y$ . The welfare gain due to personalizing recommendations is measured as  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n) = \mathbb{E}[\max_y Z_y] - (1 - \rho)\mathbb{E}[q_{(n:n)}] - \rho\mu_\varphi$ . Whether this welfare gain is large depends on which distribution,  $P_q$  or  $P_\varphi$ , has the heavier Pareto tail and on the level of heterogeneity  $\rho$ . If  $\alpha_q \neq \alpha_\varphi$ , whichever is heavier *dominates* the highest potential utility. When the exponents match, *both* matter; if  $\rho$  is small, the common term drives utility, yielding minimal additional benefit from personalization. Conversely, if  $\rho$  is large, idiosyncratic term drives utility, making personalized recommendations crucial.
  - Case  $\alpha_q < \alpha_\varphi$  (common term heavier): The maximum common term dominates, so revealing the idiosyncratic terms adds negligible extra value (see Theorem (9.b.i)).
  - Case  $\alpha_q > \alpha_\varphi$  (idiosyncratic term heavier): The maximum idiosyncratic term dominates, so revealing the idiosyncratic terms significantly boosts agent welfare (see Theorem (9.b.i)).
  - Case  $\alpha_q = \alpha_\varphi$  (both terms are equally heavy): Here,  $Z_y = (1 - \rho) q_y + \rho \varphi_y$  follows a combined Pareto tail that depends on  $c_q, c_\varphi, \alpha$  and  $\rho$ . The incremental gain of personalization depends on how strongly  $\rho$  weights  $\varphi_y$ . When  $\rho$  is small, personalization adds minimal value; when  $\rho$  is large, it is crucial (Figures 5.3c, 5.3d). To see this clearly,

consider the case when  $c_q = c_\varphi = c$ . We have that  $\Delta_{q \rightarrow u}^{\text{uncap}}(n) \asymp g(\rho; \alpha) \cdot C \cdot n^{1/\alpha}$ , where  $C$  depends on  $c$  and  $\alpha$  and  $g(\rho; \alpha) = ((1 - \rho)^\alpha + \rho^\alpha)^{1/\alpha} - (1 - \rho)$ . For large values of  $\alpha$ , we see in Figure 5.3b that  $g(\rho; \alpha)$  is nearly flat for small values of  $\rho$  ( $\rho \in (0, 1/2)$ ) and increases linearly for large values of  $\rho$  ( $\rho \in (1/2, 1)$ ). Note that as  $\alpha \rightarrow \infty$ , we have that  $g(\rho; \alpha) \rightarrow \max\{2\rho - 1, 0\}$  – for  $\rho \in (0, 1/2)$ , we have that  $g(\rho; \alpha) = 0$  and for  $\rho \in (1/2, 1)$ , we have that  $g(\rho; \alpha) = 2\rho - 1 > 0$ . This is the same phase transition we observe in the case of exponential-tailed distributions (discussed later; also see Appendix D.2.1 for a brief discussion).

- **Proof Sketch:** In the Only Quality Information regime, the agent’s expected utility is  $(1 - \rho)\mathbb{E}[q_{(n:n)}] + \rho\mu_q$ . In the Full Information regime, the agent’s expected utility depends on  $\alpha_q, \alpha_\varphi, c_q, c_\varphi$  and  $\rho$  as
  - Case  $\alpha_q < \alpha_\varphi$  (common term heavier): Since the common term dominates, we have that  $\mathbb{E}[\max_y Z_y] \asymp (1 - \rho)\mathbb{E}[q_{(n:n)}]$  which implies the result since  $\lim_{n \rightarrow \infty} \rho\mu_q / \mathbb{E}[q_{(n:n)}] = 0$ .
  - Case  $\alpha_q > \alpha_\varphi$  (idiosyncratic term heavier): Since the idiosyncratic term dominates, we have that  $\mathbb{E}[\max_y Z_y] \asymp \rho\mathbb{E}[\varphi_{(n:n)}]$  and  $\lim_{n \rightarrow \infty} \mathbb{E}[q_{(n:n)}] / \mathbb{E}[\varphi_{(n:n)}] = 0$ . Combining the two gives the result.
  - Case  $\alpha_q = \alpha_\varphi$  (both terms are equally heavy): We show that the random variable  $Z_y$  has a Pareto tail with parameters  $(c_Z, \alpha)$  where  $c_Z = (((1 - \rho)c_q)^\alpha + (\rho c_\varphi)^\alpha)^{1/\alpha}$ . This allows us to show that  $\mathbb{E}[\max_y Z_y] \asymp (c_Z/c_q)\mathbb{E}[q_{(n:n)}]$  and the result follows.

**Theorem 10 (Uncapacitated supply, Exponential tails)** *Consider the uncapacitated supply setting. Fix  $c_q > 0, \lambda_q > 0, c_\varphi > 0, \lambda_\varphi > 0$ . Assume that the common terms  $(q_y)$  are drawn i.i.d from a distribution  $P_q$  with non-negative support, finite mean  $\mu_q < \infty$  and an exponential tail with parameters  $(c_q, \lambda_q)$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from a distribution  $P_\varphi$  with non-negative support, finite mean  $\mu_\varphi < \infty$  and an exponential tail with parameters  $(c_\varphi, \lambda_\varphi)$ . For any  $\rho \in (0, 1)$ , we have that*

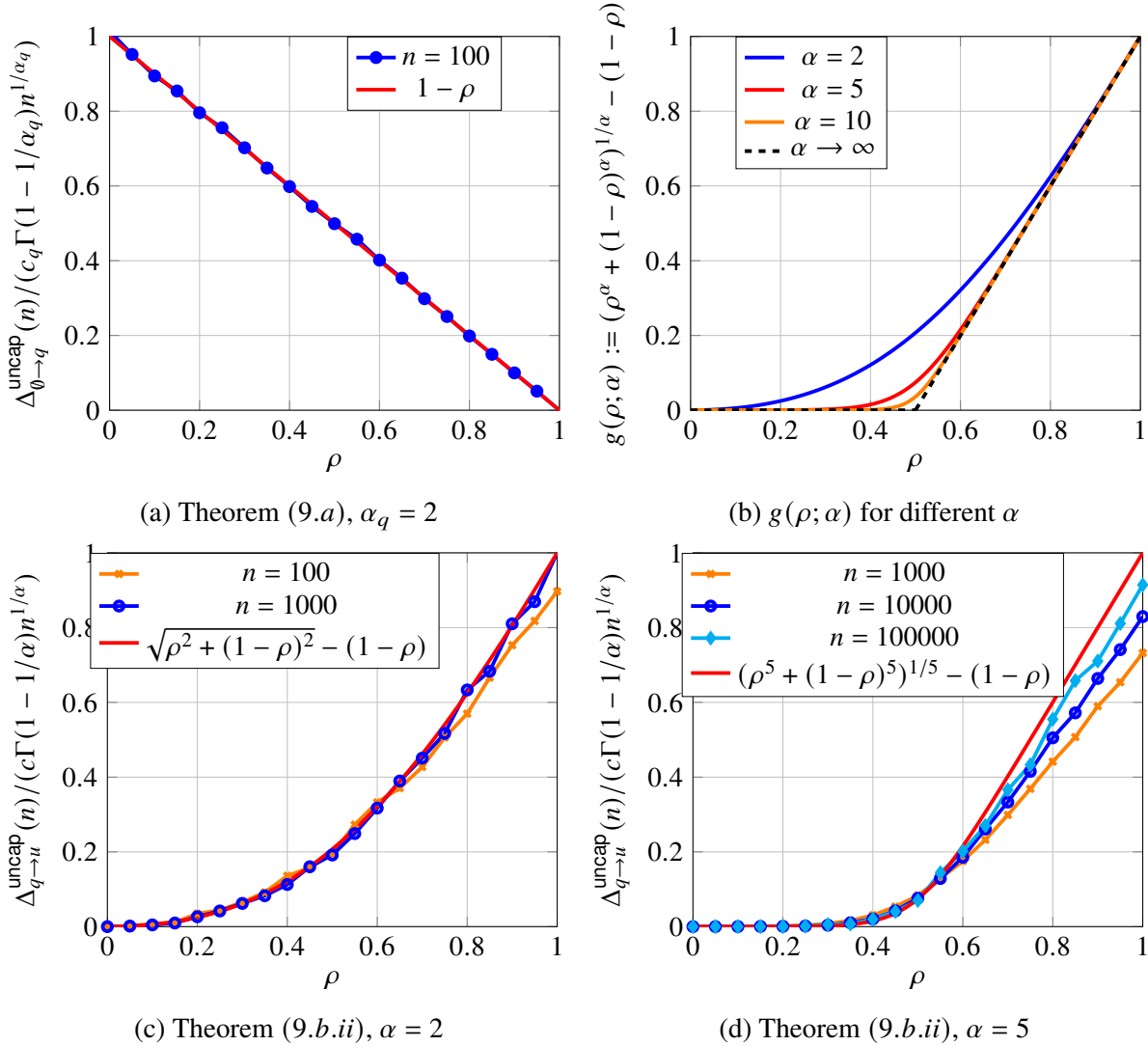


Figure 5.3: (a) Simulation plot of  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)/c_q \Gamma(1 - \alpha_q) \cdot n^{1/\alpha_q}$  as a function of  $\rho \in [0, 1]$  where  $P_q$  and  $P_\varphi$  are Pareto distributions with  $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 2$ , (b) Plot of  $g(\rho; \alpha)$  for different values of  $\alpha$ , (c) Simulation plot of  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/(\Gamma(1 - 1/\alpha)n^{1/\alpha})$  as a function of  $\rho \in [0, 1]$  where  $P_q$  and  $P_\varphi$  are Pareto distributions with  $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 2$ , (d) Simulation plot of  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/(\Gamma(1 - 1/\alpha)n^{1/\alpha})$  as a function of  $\rho \in [0, 1]$  where  $P_q$  and  $P_\varphi$  are Pareto distributions with  $c_q = c_\varphi = 1, \alpha_q = \alpha_\varphi = 5$ .

(10.a) *The difference in the agent welfare  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  obtained in the Only Quality Information regime and No Information regime increases in the number of items  $n$ . In particular, we have that*

$$\lim_{n \rightarrow \infty} \frac{\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)}{\ln n / \lambda_q} = 1 - \rho.$$

(10.b) *The difference in the agent welfare  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)$  obtained in the Full Information regime and Only Quality Information regime depends on the values of rate parameters  $\lambda_q, \lambda_\varphi$  and parameter  $\rho$ . In particular, we have that*

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\ln n} = \max \left\{ \frac{1 - \rho}{\lambda_q}, \frac{\rho}{\lambda_\varphi} \right\} - \frac{1 - \rho}{\lambda_q}.$$

*Furthermore, if  $\lambda_q = \lambda_\varphi = \lambda$ , we have that*

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\ln n / \lambda} = (2\rho - 1)_+.$$

The proof of Theorem 10 is deferred to Section D.2.3. Theorem 10 parallels our Pareto-tail results, but now the tail parameters  $\lambda_q$  and  $\lambda_\varphi$  drives the marginal welfare of public rankings and personalized recommendations.

### Discussion of Theorem (10.a): Value of Public Rankings

- **Main Insights:** In the uncapacitated setting, revealing the common term ( $q_y$ ) again yields a substantial welfare boost if  $\rho < 1$ . Specifically, Theorem (10.a) shows  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  grows asymptotically like  $\ln n / \lambda_q$ , multiplied by  $(1 - \rho)$  (see Figure 5.4a).
- **Proof Sketch:** The proof follows the same recipe as in the case of Pareto-tailed distribution with the key distinction being that maximum of common terms scales as  $\ln n / \lambda_q$ .

### Discussion of Theorem (10.b): (Incremental) Value of Personalized Recommendations

- **Main Insights:** In the Full Information regime, the agent observes both the common terms  $q_y$  and the idiosyncratic terms  $\varphi_y$  and chooses the maximum of  $n$  draws of  $Z_y = (1 - \rho)q_y + \rho\varphi_y$ . Theorem (10.b) characterizes  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)$  showing the dominant rate (either  $\lambda_q/(1 - \rho)$  or  $\lambda_\varphi/\rho$ ) determines how much extra value personalization provides.

- $\lambda_q/(1 - \rho) < \lambda_\varphi/\rho$ : There is limited benefit to revealing the idiosyncratic terms.
- $\lambda_q/(1 - \rho) > \lambda_\varphi/\rho$ : Revealing the idiosyncratic terms significantly increases utility.
- $\lambda_q = \lambda_\varphi$ : There is a knife edge transition at  $\rho = 1/2$  (see Theorem (10.b) and Figure 5.4b): for  $\rho \in (0, 1/2)$ , personalizing recommendations provide no asymptotic gain over public rankings, whereas for  $\rho \in (1/2, 1)$ , personalization offers substantial additional value.

- **Proof Sketch:** The proof follows the same recipe as in the case of Pareto-tailed distributions. The key distinction is that we show that random variable  $Z_y$  *approximately* has an exponential tail with rate  $\lambda_Z = \min\{\lambda_q/(1 - \rho), \lambda_\varphi/\rho\}$ . This result allows us characterize the scaling of  $\mathbb{E}[\max_y Z_y]$  which in turn leads to the result.

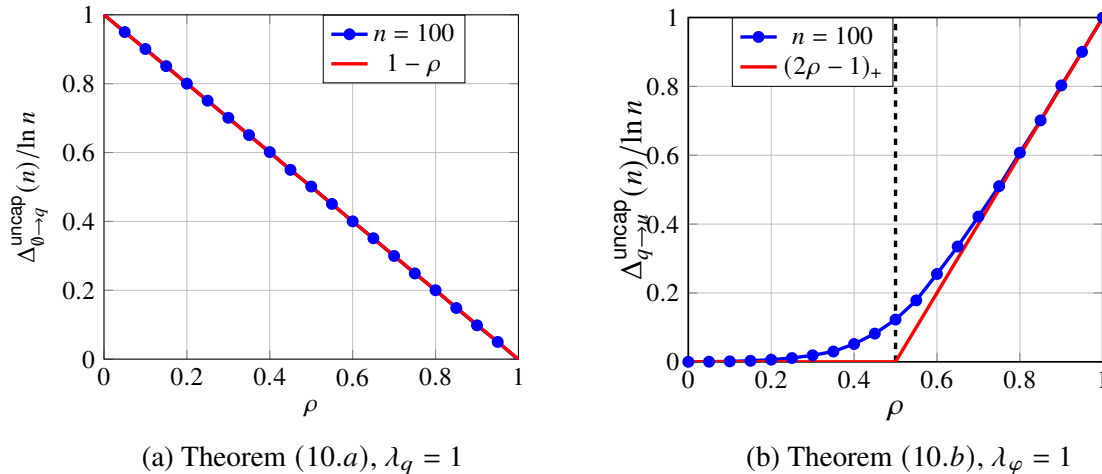


Figure 5.4: Simulation plot of (a)  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)/\ln n$  and (b)  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)/\ln n$  as a function of  $\rho \in [0, 1]$  where  $P_q$  and  $P_\varphi$  are exponential distributions with rate  $\lambda_q = \lambda_\varphi = 1$ .

### 5.3.2 Capacitated supply setting

Recall that in the capacitated supply setting, we have  $n$  agents and  $n$  items where each agent has a unit demand and each item has a unit capacity and there is one-to-one match between agents and items.

**Theorem 11 (Capacitated Supply, Pareto tails)** *Consider the capacitated supply setting. Assume that the common terms  $(q_y)$  are drawn i.i.d from distribution  $P_q$  with non-negative support and finite mean  $\mu_q < \infty$ . Fix  $c_\varphi > 0$  and  $\alpha_\varphi > 1$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from distribution  $P_\varphi$  with non-negative support, finite mean  $\mu_\varphi < \infty$  and has a Pareto tail with parameters  $(c_\varphi, \alpha_\varphi)$ . For any  $\rho \in [0, 1]$ , we have that*

(11.a) *The difference in the agent welfare  $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$  obtained in the Only Quality Information regime and the No Information regime is zero, i.e.,  $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n) = 0$ .*

(11.b) *The difference in the agent welfare  $\Delta_{q \rightarrow \varphi}^{\text{cap}}(n)$  obtained in the Full Information regime and the Only Quality Information regime increases in the number of items  $n$ . Define  $C_\varphi \triangleq c_\varphi(\alpha_\varphi/(\alpha_\varphi + 1))\Gamma(1 - 1/\alpha_\varphi)$ . Then we have that,*

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow \varphi}^{\text{cap}}(n)}{C_\varphi \cdot n^{1/\alpha_\varphi}} = \rho.$$

The proof of Theorem 11 is deferred to Section 5.4.3. Theorem 11 analyzes how public rankings and personalized recommendations affect welfare in a capacity-constrained setting.

- Theorem (11.a) studies whether Only Quality Information (public rankings) improves welfare over No Information.
- Theorem (11.b) studies the additional benefit from Only Quality Information (public rankings) to Full Information (personalized recommendations).

### Discussion of Theorem (11.a) (Value of Public Rankings)

- **Main Insights:** Even if the common terms  $q_y$  follow *any* distribution (not necessarily Pareto or exponential), public rankings do *not* increase the total welfare under unit-capacity constraints (see Figure 5.5a). The key reason is that the *total common value* across items is limited by capacity constraints, so revealing  $q_y$  merely reshuffles who claims which item but does not increase the aggregate utility. Moreover, this conclusion remains valid in a more general setting where items can have capacities  $C_y > 0$  (see Remark 7).
- **Proof Sketch:** In Only Quality Information regime, each agent bases their choice on  $q_y$ , but the idiosyncratic component  $\varphi_{xy}$  is an independent random draw. Since every agent effectively gets a “fresh” idiosyncratic draw for whichever item they pick, the expected total utility matches that in No Information. This argument requires (i) independence between  $q_y$  and  $\varphi_{xy}$  and (ii) a bounded sum of  $q_y$ ’s. See Section 5.4.3 for details.

### Discussion of Theorem (11.b) (Value of Personalized Recommendations)

- **Main Insights:** In the capacitated setting, all the value lies in personalized recommendations. Allowing agents to see both the common and idiosyncratic terms (Full Information) generates substantial welfare gains (see Figure 5.5b). We show that  $\Delta_{q \rightarrow u}^{\text{cap}}(n) \asymp \rho \cdot C_\varphi \cdot n^{1/\alpha_\varphi}$ , where  $C_\varphi$  is a constant which depends on  $c_\varphi$  and  $\alpha_\varphi$ . Revealing  $\varphi_{xy}$  matches each agent to an item that offers higher individual utility—significantly boosting total welfare. Our result also illuminates the role of level of heterogeneity. The welfare gain due to personalization of recommendations scale linearly in  $\rho$ : larger the value of  $\rho$ , larger the heterogeneity in preferences and larger the impact of personalized recommendations.
- **Proof Sketch:** By deferred decisions [159], we can imagine that when agent  $k$  arrives, the  $n - k$  relevant idiosyncratic values are drawn afresh. Thus, agent  $k$ ’s final utility is at most  $(1 - \rho) q_{\sigma_u(k)} + \rho \varphi_{(n-k:n-k)}$  and at least  $\rho \varphi_{(n-k:n-k)}$ . Summing over all agents yields a total gain on the order of  $\rho n^{1/\alpha_\varphi}$ , because the maximum of  $(n - k)$  Pareto draws scales like

$(n - k)^{1/\alpha_\varphi}$ . We formalize this argument in Lemma 4 and Proposition 4. See Section 5.4.3 for the complete proof.

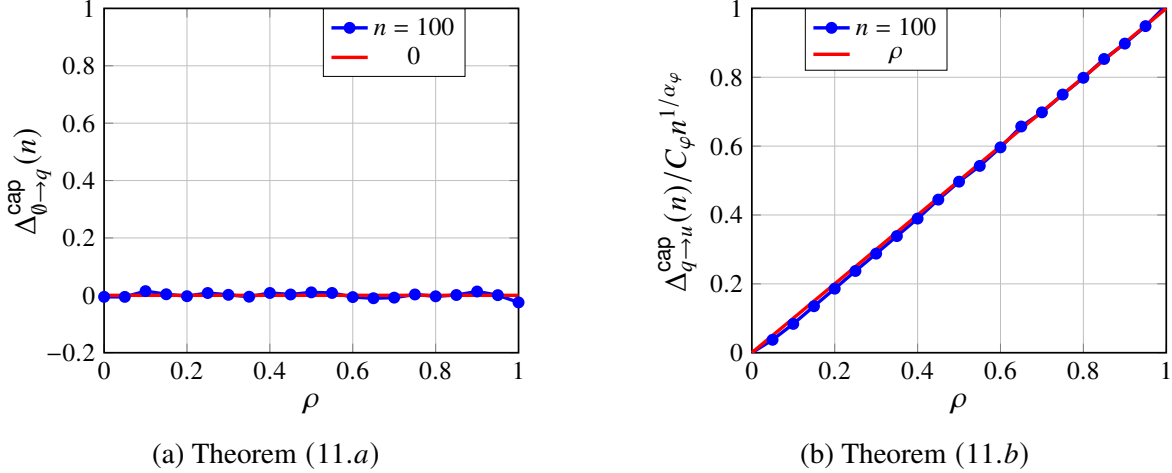


Figure 5.5: Simulation plot of  $\Delta_{\theta \rightarrow q}^{\text{cap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{cap}}(n)/C_\varphi n^{1/\alpha_\varphi}$  as a function of  $\rho \in [0, 1]$  when  $P_q$  and  $P_\varphi$  are the Pareto distribution with exponent  $\alpha_q = \alpha_\varphi = 2$  and  $c_q = c_\varphi = 1$ .

**Theorem 12 (Capacitated Supply, Exponential tails)** *Consider the capacitated supply setting. Assume that common terms  $(q_y)$  are drawn i.i.d from distribution  $P_q$  with non-negative support and finite mean  $\mu_q < \infty$ . Fix  $c_\varphi > 0, \lambda_\varphi > 0$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from distribution  $P_\varphi$  with non-negative support, finite mean  $\mu_\varphi < \infty$  and has an exponential tail with parameters  $(c_\varphi, \lambda_\varphi)$ . For any  $\rho \in [0, 1]$ , we have that*

(12.a) *The difference in the agent welfare  $\Delta_{\theta \rightarrow q}^{\text{cap}}(n)$  obtained in the Only Quality Information regime and the No Information regime is zero, i.e.,  $\Delta_{\theta \rightarrow q}^{\text{cap}}(n) = 0$ .*

(12.b) *The difference in the agent welfare  $\Delta_{q \rightarrow u}^{\text{cap}}(n)$  obtained in the Full Information regime and the Only Quality Information regime increases in the number of items  $n$ . In particular, we have that*

$$\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{cap}}(n)}{\ln n / \lambda_\varphi} = \rho.$$

We prove Theorem 12 in Appendix D.2.4. The only essential difference from Theorem 11 is the *scaling* of  $\Delta_{q \rightarrow u}^{\text{cap}}(n)$ , which here grows as  $\rho \ln(n)/\lambda_\varphi$  (rather than  $n^{1/\alpha_\varphi}$ ). This follows from the fact that the maximum of  $n$  i.i.d. exponential( $\lambda$ ) random variables scales on the order of  $\ln(n)/\lambda$  (see Proposition 6). Figures 5.6a and 5.6b illustrate Theorem (12.a) and (12.b) via numerical simulations, assuming an exponential distribution with rate  $\lambda = 1$  for the idiosyncratic terms.

**Remark 7 (Relaxing the Unit-Capacity Assumption)** *We can relax our model by allowing each item  $y \in \mathcal{Y}$  to have capacity  $C_y \in \mathbb{N}_{>0}$ , with the total number of agents equal to  $\sum_y C_y$ , i.e., a balanced market setting. Under this generalized setting, Theorems (11.a) and (12.a) remain valid, preserving the core insight that public rankings provide little value since the total common value is limited by capacity constraints. Personalized recommendations also continue to yield significant gains, though the resulting welfare expressions become more involved. Consequently, we focus on the simpler case where  $C_y = 1$ .*

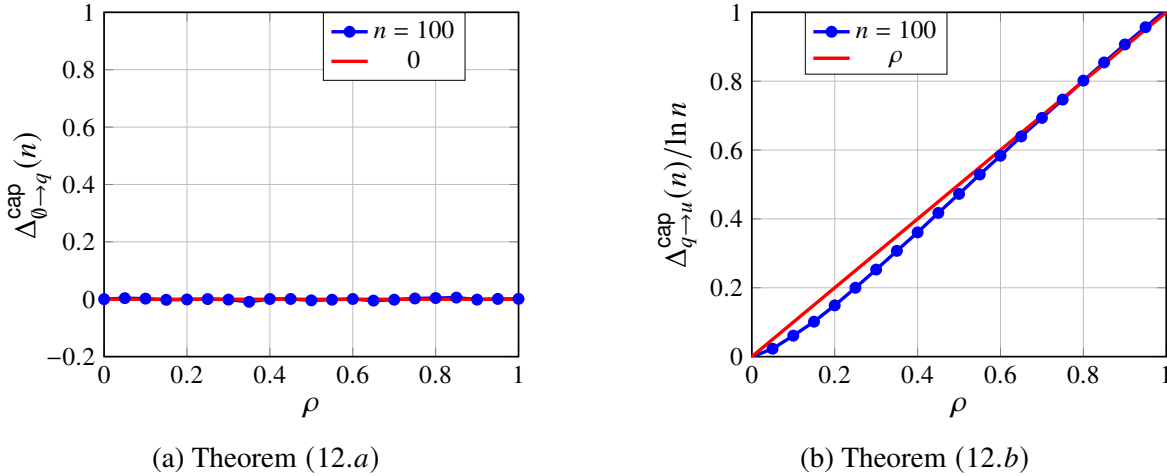


Figure 5.6: Simulation plot of  $\Delta_{\theta \rightarrow q}^{\text{cap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{cap}}(n)/\ln n$  as a function of  $\rho \in [0, 1]$  when  $P_q$  and  $P_\varphi$  are the exponential distribution with rate  $\lambda_q = \lambda_\varphi = 1$ .

#### 5.4 Proof of Theorems for utility distributions with Pareto tail

In this section, we provide the proof of our results for distributions with Pareto tails, i.e., Theorem 9 for the uncapacitated supply setting and Theorem 11 for the capacitated supply setting. The

case of distribution with exponential tails shares common ideas to that of distributions with Pareto tails and hence has been deferred to Appendix D.2.

#### 5.4.1 Useful Results

We will first state a useful proposition which will be used in proving Theorems 9 and (11.b).

**Proposition 4** *Fix  $c > 0$  and  $\alpha > 1$ . Let  $X$  be a random variable with distribution  $P$ . Assume that  $X \geq 0$  and  $\mathbb{E}[X] < \infty$ . Assume that the distribution  $P$  has a Pareto tail with parameters  $(c, \alpha)$ . Let  $X_1, X_2, \dots, X_n$  be i.i.d copies of  $X$  and define  $X_{(n:n)} \triangleq \max_{1 \leq k \leq n} X_k$ . We have that*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{c\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} = 1.$$

The proof of Proposition 4 is deferred to Appendix D.1.1.

#### 5.4.2 Uncapacitated Supply Setting

In this section, we will provide the proof of Theorem 9.

##### **Proof of Theorem (9.a)**

*Proof of Theorem (9.a).* In the uncapacitated setting, there is a single agent and  $n$  items and in the No Information regime, the agent chooses a random item since the agent has no information about the common terms ( $q_y$ ) or the idiosyncratic terms ( $\varphi_{xy}$ ). Recall that  $\sigma_\emptyset(1)$  denotes the index of the item chosen by the agent. Therefore, the social welfare  $\text{AW}_\emptyset^{\text{uncap}}(n)$  is given as

$$\text{AW}_\emptyset^{\text{uncap}}(n) = \mathbb{E} \left[ (1 - \rho)q_{\sigma_\emptyset(1)} + \rho\varphi_{1\sigma_\emptyset(1)} \right] = (1 - \rho)\mu_q + \rho\mu_\varphi, \quad (5.2)$$

where the last equality follows from the fact that index  $\sigma_\emptyset(1)$  is uniformly random in  $\{1, 2, \dots, n\}$ .

In the Only Quality Information regime, the agent chooses the item with highest common

term value. Therefore, the social welfare  $AW_q^{\text{uncap}}(n)$  is given as

$$AW_q^{\text{uncap}}(n) = \mathbb{E} \left[ (1 - \rho)q_{\sigma_q(1)} + \rho\varphi_{1\sigma_q(1)} \right] = \mathbb{E} \left[ (1 - \rho)q_{(n:n)} + \rho\varphi_{1\sigma_q(1)} \right] = (1 - \rho)\mathbb{E}[q_{(n:n)}] + \rho\mu_\varphi, \quad (5.3)$$

where the last equality follows from the fact that index  $\sigma_q(1)$  is uniformly random in  $\{1, 2, \dots, n\}$ .

Using (5.2), (5.3) and Proposition 4, we get the required result.  $\blacksquare$

### Proof of Theorem (9.b)

We will begin by stating an important result in the form of Lemma 3 which will be crucial in proving Theorem (9.b).

**Lemma 3** *Fix  $\rho \in (0, 1)$ . Fix  $c_X > 0, \alpha_X > 1, c_Y > 0, \alpha_Y > 1$ . Let  $X$  be a random variable with non-negative support, finite mean  $\mu_X < \infty$  and has a Pareto tail with parameters  $(c_X, \alpha_X)$ . Let  $Y$  be another random variable with non-negative support, finite mean  $\mu_Y < \infty$  and has a Pareto tail with parameters  $(c_Y, \alpha_Y)$ . Define  $Z = (1 - \rho)X + \rho Y$ . Then we have that  $Z \geq 0$ ,  $\mathbb{E}[Z] = \mu_Z = (1 - \rho)\mu_X + \rho\mu_Y < \infty$  and has a Pareto tail with parameters  $(c_Z, \alpha_Z)$  given as*

$$(3.a) \quad \underline{\alpha_X < \alpha_Y}. \quad c_Z = (1 - \rho)c_X \text{ and } \alpha_Z = \alpha_X.$$

$$(3.b) \quad \underline{\alpha_X > \alpha_Y}. \quad c_Z = \rho c_Y \text{ and } \alpha_Z = \alpha_Y.$$

$$(3.c) \quad \underline{\alpha_X = \alpha_Y := \alpha}. \quad c_Z = (((1 - \rho)c_X)^\alpha + (\rho c_Y)^\alpha)^{1/\alpha} \text{ and } \alpha_Z = \alpha.$$

Lemma 3 follows from [160, Lemma 2.18] and for completeness we provide a proof in Appendix D.1.2.

*Proof of Theorem (9.b).* Since there is only one agent, we will denote  $\varphi_{1k} = \varphi_k$  for all  $k \in \{1, 2, \dots, n\}$ . We will begin by proving part (9.b.i). Define  $Z_k = (1 - \rho)q_k + \rho\varphi_k$ . In the Full Information regime, the agent will choose the item with the value  $\max\{Z_1, Z_2, \dots, Z_n\}$ . Therefore,

the social welfare  $AW_u^{\text{uncap}}(n)$  is given as

$$AW_u^{\text{uncap}}(n) = \mathbb{E} \left[ \max_{1 \leq k \leq n} (1 - \rho)q_k + \rho\varphi_k \right] = \mathbb{E}[Z_{(n:n)}]. \quad (5.4)$$

Next we consider the different cases:

(i)  $\alpha_q \neq \alpha_\varphi$ . Assume that  $\alpha_q < \alpha_\varphi$ , then we have that

$$\begin{aligned} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} &\stackrel{(a)}{=} \frac{\mathbb{E}[Z_{(n:n)}]}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} - \frac{(1 - \rho)\mathbb{E}[q_{(n:n)}]}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} - \frac{\rho\mu_\varphi}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}}, \\ &\stackrel{(b)}{=} \frac{(1 - \rho)\mathbb{E}[Z_{(n:n)}]}{c_Z \Gamma(1 - 1/\alpha_Z) n^{1/\alpha_Z}} - \frac{(1 - \rho)\mathbb{E}[q_{(n:n)}]}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} - \frac{\rho\mu_\varphi}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}}, \end{aligned}$$

where (a) follows from the fact that  $\Delta_{q \rightarrow u}^{\text{uncap}}(n) = AW_u^{\text{uncap}}(n) - AW_q^{\text{uncap}}(n)$  and (5.3) and (5.4), (b) follows from Lemma 3 for  $\alpha_q < \alpha_\varphi$ . Using Proposition 4, we have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{c_Z \Gamma(1 - 1/\alpha_Z) n^{1/\alpha_Z}} = 1, \quad \lim_{n \rightarrow \infty} \frac{\mathbb{E}[q_{(n:n)}]}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} = 1,$$

which in turn implies that  $\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c_q \Gamma(1 - 1/\alpha_q) n^{1/\alpha_q}} = 0$ .

Next we assume that  $\alpha_q > \alpha_\varphi$ , then we have that

$$\begin{aligned} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}} &\stackrel{(a)}{=} \frac{\mathbb{E}[Z_{(n:n)}]}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}} - \frac{(1 - \rho)\mathbb{E}[q_{(n:n)}]}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}} - \frac{\rho\mu_\varphi}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}}, \\ &\stackrel{(b)}{=} \frac{\rho\mathbb{E}[Z_{(n:n)}]}{c_Z \Gamma(1 - 1/\alpha_Z) n^{1/\alpha_Z}} - \frac{(1 - \rho)\mathbb{E}[q_{(n:n)}]}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}} - \frac{\rho\mu_\varphi}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}}, \end{aligned}$$

where (a) follows from the fact that  $\Delta_{q \rightarrow u}^{\text{uncap}}(n) = AW_u^{\text{uncap}}(n) - AW_q^{\text{uncap}}(n)$  and (5.3) and (5.4), (b) follows from Lemma 3 for  $\alpha_q > \alpha_\varphi$ . Using Proposition 4, since  $\alpha_q > \alpha_\varphi$ , we have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[q_{(n:n)}]}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_\varphi}} = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[q_{(n:n)}]}{c_\varphi \Gamma(1 - 1/\alpha_\varphi) n^{1/\alpha_q}} \cdot \lim_{n \rightarrow \infty} \frac{n^{1/\alpha_q}}{n^{1/\alpha_\varphi}} = 0,$$

which in turn implies that  $\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{c_\varphi \Gamma(1-1/\alpha_\varphi) n^{1/\alpha_\varphi}} = \rho$ .

(ii)  $\alpha_q = \alpha_\varphi$ . Denote  $\alpha := \alpha_q = \alpha_\varphi$ . Then we have that,

$$\frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\Gamma(1-1/\alpha) n^{1/\alpha}} \stackrel{(a)}{=} \frac{\mathbb{E}[Z_{(n:n)}]}{\Gamma(1-1/\alpha) n^{1/\alpha}} - \frac{(1-\rho)\mathbb{E}[q_{(n:n)}]}{\Gamma(1-1/\alpha) n^{1/\alpha}} - \frac{\rho\mu_\varphi}{\Gamma(1-1/\alpha) n^{1/\alpha}},$$

where (a) follows from the fact that  $\Delta_{q \rightarrow u}^{\text{uncap}}(n) = \text{AW}_u^{\text{uncap}}(n) - \text{AW}_q^{\text{uncap}}(n)$  and (5.3) and (5.4). Using Lemma 3 and Proposition 4, we have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{\Gamma(1-1/\alpha) n^{1/\alpha}} = (((1-\rho)c_q)^\alpha + (\rho c_\varphi)^\alpha)^{1/\alpha}, \quad \lim_{n \rightarrow \infty} \frac{\mathbb{E}[q_{(n:n)}]}{\Gamma(1-1/\alpha) n^{1/\alpha}} = (1-\rho)c_q,$$

which in turn implies that  $\lim_{n \rightarrow \infty} \frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\Gamma(1-1/\alpha) n^{1/\alpha}} = (((1-\rho)c_q)^\alpha + (\rho c_\varphi)^\alpha)^{1/\alpha} - (1-\rho)c_q$ .

The case of  $c_q = c_\varphi = c$  follows trivially.

This completes the proof. ■

### 5.4.3 Capacitated supply setting

In this section, we will provide the proof of Theorem 11. Theorem 11 has two parts: (a) characterizes the difference between the Only Quality Information regime and No Information regime  $\Delta_{\varphi \rightarrow q}^{\text{cap}}(n)$  and (b) characterizes the difference between the Full Information regime and the Only Quality Information regime  $\Delta_{q \rightarrow u}^{\text{cap}}(n)$ .

#### **Proof of Theorem (11.a)**

*Proof of Theorem (11.a).* In the No Information regime, since the agents do not have information about the common term or the idiosyncratic term, they randomly choose an item from the remaining set of items. Recall that  $\sigma_\theta(k)$  denotes the index of the item chosen by agent  $k$  in the

No Information regime. Therefore we have that,

$$\text{AW}_\emptyset^{\text{cap}}(n) \stackrel{(a)}{=} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n (1 - \rho) q_{\sigma_\emptyset(k)} + \rho \varphi_{k\sigma_\emptyset(k)} \right] \stackrel{(b)}{=} (1 - \rho) \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n q_k \right] + \rho \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\varphi_{k\sigma_\emptyset(k)}], \quad (5.5)$$

where (a) follows from definition of  $u_{k\sigma(k)}$ , (b) follows from the fact that  $\sum_{k=1}^n q_{\sigma_\emptyset(k)} = \sum_{k=1}^n q_k$ .

In the Only Quality Information regime, the agents base their decisions solely on the common term ( $q_y$ ). Therefore, we have that the agent  $k$  will choose the item with common term value  $q_{(k:n)}$  (recall that  $X_{(k:n)}$  denotes the  $k$ -th smallest value of  $n$  i.i.d copies of  $X$ ). Recall that  $\sigma_q(k)$  denotes the index of the item chosen by agent  $k$  in the Only Quality Information regime. This means that  $q_{\sigma_q(k)} = q_{(k:n)}$ . Therefore we have that,

$$\text{AW}_q^{\text{cap}}(n) \stackrel{(a)}{=} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n (1 - \rho) q_{\sigma_q(k)} + \rho \varphi_{k\sigma_q(k)} \right] \stackrel{(b)}{=} (1 - \rho) \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n q_k \right] + \rho \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\varphi_{k\sigma_q(k)}], \quad (5.6)$$

where (a) follows from definition of  $u_{k\sigma_q(k)}$ , (b) follows from the fact that  $\sum_{k=1}^n q_{\sigma_\emptyset(k)} = \sum_{k=1}^n q_{(k:n)} = \sum_{k=1}^n q_k$ . Note that the index  $\sigma_\emptyset(k)$  and  $\sigma_q(k)$  are random and hence we have that  $\varphi_{k\sigma_\emptyset(k)} \stackrel{d}{=} \varphi_{k\sigma_q(k)}$  (have the same distribution) and therefore  $\mathbb{E}[\varphi_{k\sigma_\emptyset(k)}] = \mathbb{E}[\varphi_{k\sigma_q(k)}]$ . Comparing (5.5) and (5.6), we have that  $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n) = 0$ . ■

### Proof of Theorem (11.b)

We first present a key lemma which will be useful in proving Theorem (11.b).

**Lemma 4** *Consider the capacitated supply setting. Assume that the common terms ( $q_y$ ) are drawn i.i.d from distribution  $P_q$  with non-negative support and finite mean  $\mu_q < \infty$ . Assume that the idiosyncratic terms ( $\varphi_{xy}$ ) are drawn i.i.d from distribution  $P_\varphi$  with non-negative support and finite mean  $\mu_\varphi$ . Let  $\varphi_{k,(n-k:n-k)}$  denote that the maximum value amongst  $n - k$  i.i.d draws from distribu-*

tion  $P_\varphi$ . Define  $\Phi_n \triangleq n^{-1} \sum_{k=1}^n \mathbb{E}[\varphi_{k,(n-k:n-k)}]$ . Then for all  $\rho \in [0, 1]$ , we have that

$$-\frac{(1-\rho)\mu_q + \rho\mu_\varphi}{\Phi_n} + \rho \leq \frac{\Delta_{q \rightarrow u}^{\text{cap}}(n)}{\Phi_n} \leq -\frac{\rho\mu_\varphi}{\Phi_n} + \rho.$$

We defer the proof of Lemma 4 to Appendix D.1.3. In the case of Theorem (11.b), it suffices to show that  $\lim_{n \rightarrow \infty} \Phi_n / (C_\varphi \cdot n^{1/\alpha_\varphi}) = 1$ , where  $C_\varphi$  is defined in Theorem (11.b).

*Proof of Theorem (11.b).* Let us denote  $\varphi_{k,(n-k:n-k)} := \varphi_{(n-k,n-k)}$ . Fix  $\epsilon > 0$ . There exists an  $k_0 \in \mathbb{N}$  such for all  $k \geq k_0$ , we have that

$$(1-\epsilon)c\Gamma(1-1/\alpha_\varphi) \cdot k^{1/\alpha_\varphi} \leq \mathbb{E}[\varphi_{(k:k)}] \leq (1+\epsilon)c\Gamma(1-1/\alpha_\varphi) \cdot k^{1/\alpha_\varphi} \quad (5.7)$$

We can upper bound  $\Phi_n$  as follows:

$$\begin{aligned} \Phi_n &\stackrel{(a)}{=} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\varphi_{(k:k)}] \stackrel{(b)}{=} \frac{1}{n} \sum_{k=1}^{m_0} \mathbb{E}[\varphi_{(k:k)}] + \frac{1}{n} \sum_{k=m_0+1}^n \mathbb{E}[\varphi_{(k:k)}], \\ &\stackrel{(c)}{\leq} \mu_\varphi \frac{m_0(m_0+1)}{2n} + \frac{1}{n} (1+\epsilon)c\Gamma(1-1/\alpha_\varphi) \sum_{k=m_0+1}^n k^{1/\alpha_\varphi}, \\ &\stackrel{(d)}{\leq} \mu_\varphi \frac{m_0(m_0+1)}{2n} + \frac{1}{n} (1+\epsilon)c\Gamma(1-1/\alpha_\varphi) \int_0^n x^{1/\alpha_\varphi} dx, \\ &\stackrel{(e)}{=} \mu_\varphi \frac{m_0(m_0+1)}{2n} + (1+\epsilon)c \frac{\alpha_\varphi}{1+\alpha_\varphi} \Gamma(1-1/\alpha_\varphi) n^{1/\alpha_\varphi}, \end{aligned}$$

where (a) follows from the definition of  $\Phi_n$ , (b) follows trivially, (c) follows from (5.7) and the fact that  $\mathbb{E}[\varphi_{(k:k)}] \leq k\mu_\varphi$  for all  $k \leq m_0$  since  $\mathbb{E}[\max\{X_1, X_2, \dots, X_k\}] \leq \mathbb{E}[\sum_{j=1}^k X_j] = k\mathbb{E}[X]$ , (d) follows from the fact that  $\sum_{k=m_0+1}^n k^{1/\alpha_\varphi} \leq \int_0^n x^{1/\alpha_\varphi} dx$ , (e) follows from evaluating the integral.

Using this we have that

$$\limsup_{n \rightarrow \infty} \frac{\Phi_n}{C_\varphi \cdot n^{1/\alpha_\varphi}} \leq 1 + \epsilon.$$

Using similar arguments as above, we can easily show that

$$\liminf_{n \rightarrow \infty} \frac{\Phi_n}{C_\varphi \cdot n^{1/\alpha_\varphi}} \geq 1 - \epsilon.$$

Since this holds for all  $\epsilon > 0$ , we have that  $\lim_{n \rightarrow \infty} \frac{\Phi_n}{C_\varphi \cdot n^{1/\alpha_\varphi}} = 1$  and this completes the proof. ■

## 5.5 Conclusion

In this work, we examine the impact of public rankings and personalized recommendations on agent welfare in different marketplace settings. To isolate and quantify the impact of these information provisioning tools, we study a stylized model where the agents utility for the items comprises of two terms: (i) a common term and (ii) an idiosyncratic term and both these terms are independent of each other. Public rankings enable the agents to learn about the common term whereas personalized recommendations help the agents to learn about their idiosyncratic component about the items. We quantify the agent welfare under different distributional assumptions on the common and the idiosyncratic terms and under different marketplace settings. Our findings reveal a fundamental interplay between the benefits of these information tools and supply-side constraints. Specifically, in supply-constrained settings, public rankings alone offer limited value in enhancing agent welfare. However, personalized recommendations unlock substantial value by refining individual utility estimates and improving the allocation of agents to items, thereby reducing congestion. Conversely, in supply-unconstrained settings, public rankings significantly enhance welfare by identifying the best overall options, while the impact of personalized recommendations becomes more nuanced. This contrast arises because public rankings primarily serve to highlight the top items in general, while personalized recommendations serve a dual role: (i) they help agents refine their utility assessments beyond what rankings provide, and (ii) they facilitate a more efficient allocation by mitigating congestion. In capacity-constrained environments, both effects of personalized recommendations are crucial, thus unlocking significant value. In environments without capacity constraints, only the first effect is relevant, leading to a situation

where both public rankings and personalized recommendations contribute, but in distinct ways, to agent welfare.

This work takes a first step toward a principled understanding of how various information-provisioning tools perform across different marketplace settings. Our model is deliberately stylized to provide crisp insights, yet it opens several avenues for further investigation. A central assumption in our analysis is the independence of the idiosyncratic terms across agent–item pairs, which plays a critical role in driving our results and simplifies significant technical challenges. In reality, these terms may be correlated, and understanding how such correlation affects the impact of different information-provisioning tools is a promising direction for future research. We have focused exclusively on *agent welfare*; extending the analysis to encompass broader objectives, such as *social welfare* (which also accounts for the utility of the supply side), would provide a more comprehensive assessment of these tools.

## References

- [1] J.-P. Aubin and I. Ekeland, “Estimates of the duality gap in nonconvex optimization,” *Mathematics of Operations Research*, vol. 1, no. 3, pp. 225–245, 1976.
- [2] D. P. Bertsekas and N. R. Sandell, “Estimates of the duality gap for large-scale separable nonconvex optimization problems,” in *1982 21st IEEE conference on decision and control*, IEEE, 1982, pp. 782–785.
- [3] M. D. Ekstrand and M. C. Willemsen, “Behaviorism is not enough: Better recommendations through listening to users,” in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 221–224.
- [4] S. Milli, L. Belli, and M. Hardt, “From optimizing engagement to measuring value,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 714–722.
- [5] Y. Chen, Y. Kanoria, A. Kumar, and W. Zhang, “Feature based dynamic matching,” *Available at SSRN 4451799*, 2023.
- [6] K. T. Talluri, G. Van Ryzin, and G. Van Ryzin, *The theory and practice of revenue management*. Springer, 2004, vol. 1.
- [7] G. Bitran and R. Caldentey, “An overview of pricing models for revenue management,” *Manufacturing & Service Operations Management*, vol. 5, no. 3, pp. 203–229, 2003.
- [8] M. Braverman, M. Derakhshan, and A. M. Lovett, “Max-weight online stochastic matching: Improved approximations against the online benchmark,” *arXiv preprint arXiv:2206.01270*, 2022.
- [9] S. Banerjee, D. Freund, and T. Lykouris, “Pricing and optimization in shared vehicle systems: An approximation framework,” *Operations Research*, vol. 70, no. 3, pp. 1783–1805, 2022.
- [10] Y. Kanoria, “Dynamic spatial matching,” in *Proceedings of the 23rd ACM Conference on Economics and Computation*, 2022, pp. 63–64.
- [11] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 1.
- [12] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer science & business media, 2013.

- [13] L. Chen, R. Kyng, Y. P. Liu, R. Peng, M. P. Gutenberg, and S. Sachdeva, “Maximum flow and minimum-cost flow in almost-linear time,” *arXiv preprint arXiv:2203.00671*, 2022.
- [14] M. Akbarpour, Y. Alimohammadi, S. Li, and A. Saberi, “The value of excess supply in spatial matching markets,” *arXiv preprint arXiv:2104.03219*, 2021.
- [15] E. Balkanski, Y. Faenza, and N. Perivier, “The power of greedy for online minimum cost matching on the line,” *arXiv preprint arXiv:2210.03166*, 2022.
- [16] A. Aouad and W. Ma, “A nonparametric framework for online stochastic matching with correlated arrivals,” *arXiv preprint arXiv:2208.02229*, 2022.
- [17] S. Delong, A. Farhadi, R. Niazadeh, and B. Sivan, “Online bipartite matching with reusable resources,” in *Proceedings of the 23rd ACM Conference on Economics and Computation*, 2022, pp. 962–963.
- [18] R. Udvani, “Periodic reranking for online matching of reusable resources,” *arXiv preprint arXiv:2110.02400*, 2021.
- [19] C. Papadimitriou, T. Pollner, A. Saberi, and D. Wajc, “Online stochastic max-weight bipartite matching: Beyond prophet inequalities,” in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021, pp. 763–764.
- [20] T. Ezra, M. Feldman, N. Gravin, and Z. G. Tang, “Online stochastic max-weight matching: Prophet inequality for vertex and edge arrival models,” in *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 769–787.
- [21] W. Ma and D. Simchi-Levi, “Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios,” *Operations Research*, vol. 68, no. 6, pp. 1787–1803, 2020.
- [22] I. Ashlagi, M. Burq, C. Dutta, P. Jaillet, A. Saberi, and C. Sholley, “Edge weighted online windowed matching,” in *Proceedings of the 2019 ACM Conference on Economics and Computation*, 2019, pp. 729–742.
- [23] A. Aouad and D. Saban, “Online assortment optimization for two-sided matching platforms,” *Management Science*, 2022.
- [24] I. Ashlagi, A. K. Krishnaswamy, R. Makhijani, D. Saban, and K. Shiragur, “Assortment planning for two-sided sequential matching markets,” *Operations Research*, vol. 70, no. 5, pp. 2784–2803, 2022.
- [25] P. Shi, “Optimal matchmaking strategy in two-sided marketplaces,” *Management Science*, 2022.

- [26] V. Manshadi, S. Rodilitz, D. Saban, and A. Suresh, “Online algorithms for matching platforms with multi-channel traffic,” *arXiv preprint arXiv:2203.15037*, 2022.
- [27] M. Derakhshan, N. Golrezaei, V. Manshadi, and V. Mirrokni, “Product ranking on online platforms,” *Management Science*, vol. 68, no. 6, pp. 4024–4041, 2022.
- [28] Y. Feng and R. Niazadeh, “Batching and optimal multi-stage bipartite allocations,” *arXiv preprint arXiv:2211.16581*, 2022.
- [29] N. Immorlica, B. Lucier, V. Manshadi, and A. Wei, “Designing approximately optimal search on matching platforms,” in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021, pp. 632–633.
- [30] A. Aouad and Ö. Saritaç, “Dynamic stochastic matching under limited time,” in *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 789–790.
- [31] C. Derman, G. J. Lieberman, and S. M. Ross, “A sequential stochastic assignment problem,” *Management Science*, vol. 18, no. 7, pp. 349–355, 1972.
- [32] S. C. Albright Jr, “Stochastic sequential assignment problems.,” STANFORD UNIV CALIF DEPT OF OPERATIONS RESEARCH, Tech. Rep., 1972.
- [33] X. Su and S. A. Zenios, “Patient choice in kidney allocation: A sequential stochastic assignment model,” *Operations research*, vol. 53, no. 3, pp. 443–455, 2005.
- [34] S. R. Balseiro, O. Besbes, and D. Pizarro, “Survey of dynamic resource-constrained reward collection problems: Unified model and analysis,” *Operations Research*, 2023.
- [35] B. Taşkesen, S. Shafieezadeh-Abadeh, and D. Kuhn, “Semi-discrete optimal transport: Hardness, regularization and numerical solution,” *Mathematical Programming*, pp. 1–74, 2022.
- [36] A. Vera and S. Banerjee, “The bayesian prophet: A low-regret framework for online decision making,” *Management Science*, vol. 67, no. 3, pp. 1368–1391, 2021.
- [37] P. Bumpensanti and H. Wang, “A re-solving heuristic with uniformly bounded loss for network revenue management,” *Management Science*, vol. 66, no. 7, pp. 2993–3009, 2020.
- [38] K. Talluri and G. Van Ryzin, “A randomized linear programming method for computing network bid prices,” *Transportation science*, vol. 33, no. 2, pp. 207–216, 1999.
- [39] S. Kunnumkal, K. Talluri, and H. Topaloglu, “A randomized linear programming method for network revenue management with product-specific no-shows,” *Transportation Science*, vol. 46, no. 1, pp. 90–108, 2012.

- [40] D. Freund and S. Banerjee, “Good prophets know when the end is near,” *Available at SSRN 3479189*, 2019.
- [41] S. R. Sinclair *et al.*, “Hindsight learning for mdps with exogenous inputs,” in *International Conference on Machine Learning*, PMLR, 2023, pp. 31 877–31 914.
- [42] N. Ahani, P. Gözl, A. D. Procaccia, A. Teytelboym, and A. C. Trapp, “Dynamic placement in refugee resettlement,” *Operations Research*, 2023.
- [43] O. Besbes, Y. Kanoria, and A. Kumar, “Dynamic resource allocation: Algorithmic design principles and spectrum of achievable performances,” *arXiv preprint arXiv:2205.09078*, 2023.
- [44] J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan, “Online stochastic matching: Beating  $1-1/e$ ,” in *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, 2009, pp. 117–126.
- [45] V. H. Manshadi, S. O. Gharan, and A. Saberi, “Online stochastic matching: Online actions based on offline statistics,” *Mathematics of Operations Research*, vol. 37, no. 4, pp. 559–573, 2012.
- [46] P. Jaillet and X. Lu, “Online stochastic matching: New algorithms with better bounds,” *Mathematics of Operations Research*, vol. 39, no. 3, pp. 624–646, 2014.
- [47] B. Brubach, K. A. Sankararaman, A. Srinivasan, and P. Xu, “New algorithms, better bounds, and a novel model for online stochastic matching,” in *24th Annual European Symposium on Algorithms (ESA 2016)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.
- [48] B. Haeupler, V. S. Mirrokni, and M. Zadimoghaddam, “Online stochastic weighted matching: Improved approximation algorithms,” in *Internet and Network Economics: 7th International Workshop, WINE 2011, Singapore, December 11-14, 2011. Proceedings 7*, Springer, 2011, pp. 170–181.
- [49] A. Saberi, M. Yang, and S. H. Yu, “Stochastic online metric matching: Adversarial is no harder than stochastic,” *arXiv preprint arXiv:2407.14785*, 2024.
- [50] N. Holden, Y. Peres, and A. Zhai, “Gravitational allocation for uniform points on the sphere,” *The Annals of Probability*, vol. 49, no. 1, pp. 287–321, 2021.
- [51] A. Gupta, G. Guruganesh, B. Peng, and D. Wajc, “Stochastic online metric matching,” *arXiv preprint arXiv:1904.09284*, 2019.
- [52] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, “Scaling hypothesis for the euclidean bipartite matching problem,” *Physical Review E*, vol. 90, no. 1, p. 012 118, 2014.

- [53] T. Manole and J. Niles-Weed, “Sharp convergence rates for empirical optimal transport with smooth costs,” *arXiv preprint arXiv:2106.13181*, 2021.
- [54] O. Besbes, F. Castro, and I. Lobel, “Spatial capacity planning,” *Operations Research*, vol. 70, no. 2, pp. 1271–1291, 2022.
- [55] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [56] C. C. Aggarwal *et al.*, *Recommender systems*. Springer, 2016, vol. 1.
- [57] J. Niles-Weed and P. Rigollet, “Estimation of wasserstein distances in the spiked transport model,” *arXiv preprint arXiv:1909.07513*, 2019.
- [58] M. Ledoux, “On optimal matching of gaussian samples,” *Journal of Mathematical Sciences*, vol. 238, no. 4, pp. 495–522, 2019.
- [59] J. Weed and F. Bach, “Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance,” *Bernoulli*, vol. 25, no. 4A, pp. 2620–2648, 2019.
- [60] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman, “Plugin estimation of smooth optimal transport maps,” *arXiv preprint arXiv:2107.12364*, 2021.
- [61] X.-N. Ma, N. S. Trudinger, and X.-J. Wang, “Regularity of potential functions of the optimal transportation problem,” *Archive for rational mechanics and analysis*, vol. 177, pp. 151–183, 2005.
- [62] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022.
- [63] Y. Brenier, “Polar factorization and monotone rearrangement of vector-valued functions,” *Communications on pure and applied mathematics*, vol. 44, no. 4, pp. 375–417, 1991.
- [64] C. Villani, *Optimal transport: old and new*. Springer, 2009, vol. 338.
- [65] O. Besbes and D. Sauré, “Dynamic pricing strategies in the presence of demand shifts,” *Manufacturing & Service Operations Management*, vol. 16, no. 4, pp. 513–528, 2014.
- [66] S. Balseiro, C. Kroer, and R. Kumar, “Online resource allocation under horizon uncertainty,” in *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 2023, pp. 63–64.
- [67] Y. Bai, O. El Housni, B. Jin, P. Rusmevichientong, H. Topaloglu, and D. P. Williamson, “Fluid approximations for revenue management under high-variance demand,” *Management Science*, 2023.

- [68] O. Besbes, Y. Kanoria, and A. Kumar, “Dynamic resource allocation: Algorithmic design principles and spectrum of achievable performances,” *Operations Research*, 2024.
- [69] A. Arlotto and I. Gurvich, “Uniformly bounded regret in the multiselection problem,” *Stochastic Systems*, vol. 9, no. 3, pp. 231–260, 2019.
- [70] R. L. Bray, “Logarithmic regret in multiselection and online linear programming problems with continuous valuations,” *arXiv e-prints*, arXiv:1912, 2022.
- [71] K. T. Talluri and G. J. Van Ryzin, *The theory and practice of revenue management*. Springer Science & Business Media, 2006, vol. 68.
- [72] S. Jasin and A. Sinha, “An lp-based correlated rounding scheme for multi-item ecommerce order fulfillment,” *Operations Research*, vol. 63, no. 6, pp. 1336–1351, 2015.
- [73] G. S. Lueker, “Average-case analysis of off-line and on-line knapsack problems,” *Journal of Algorithms*, vol. 29, no. 2, pp. 277–305, 1998.
- [74] A. Cayley, “Mathematical questions with their solutions,” *The Educational Times*, vol. 23, pp. 18–19, 1875.
- [75] L. Moser, “On a problem of cayley,” *Scripta Math*, vol. 22, pp. 289–292, 1956.
- [76] A. J. Kleywegt and J. D. Papastavrou, “The dynamic and stochastic knapsack problem,” *Operations research*, vol. 46, no. 1, pp. 17–35, 1998.
- [77] R. Kleinberg, “A multiple-choice secretary algorithm with applications to online auctions,” in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, Citeseer, 2005, pp. 630–631.
- [78] L. Blumrosen and T. Holenstein, “Posted prices vs. negotiations: An asymptotic analysis.” *EC*, vol. 10, pp. 1 386 790–1 386 801, 2008.
- [79] A. Vera, S. Banerjee, and I. Gurvich, “Online allocation and pricing: Constant regret via bellman inequalities,” *Operations Research*, 2021.
- [80] S. R. Sinclair, F. Frujeri, C.-A. Cheng, and A. Swaminathan, “Hindsight learning for mdps with exogenous inputs,” *arXiv preprint arXiv:2207.06272*, 2022.
- [81] M. T. Hajiaghayi, R. Kleinberg, and T. Sandholm, “Automated online mechanism design and prophet inequalities,” in *AAAI*, vol. 7, 2007, pp. 58–65.
- [82] S. Alaei, “Bayesian combinatorial auctions: Expanding single buyer mechanisms to many buyers,” *SIAM Journal on Computing*, vol. 43, no. 2, pp. 930–972, 2014.

- [83] S. Chawla, N. Devanur, and T. Lykouris, “Static pricing for multi-unit prophet inequalities,” *arXiv preprint arXiv:2007.07990*, 2020.
- [84] J. Jiang, W. Ma, and J. Zhang, “Tight guarantees for multi-unit prophet inequalities and online stochastic knapsack,” in *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SIAM, 2022, pp. 1221–1246.
- [85] A. Arlotto and X. Xie, “Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards,” *Stochastic Systems*, vol. 10, no. 2, pp. 170–191, 2020.
- [86] X. Li and Y. Ye, “Online linear programming: Dual convergence, new algorithms, and regret bounds,” *Operations Research*, vol. 70, no. 5, pp. 2948–2966, 2022.
- [87] J. Jiang, W. Ma, and J. Zhang, “Degeneracy is ok: Logarithmic regret for network revenue management with indiscrete distributions,” *arXiv preprint arXiv:2210.07996*, 2022.
- [88] K. Talluri and G. Van Ryzin, “An analysis of bid-price controls for network revenue management,” *Management science*, vol. 44, no. 11-part-1, pp. 1577–1593, 1998.
- [89] X. Shen and S. Boyd, “Incremental proximal multi-forecast model predictive control,” *arXiv preprint arXiv:2111.14728*, 2021.
- [90] S. Balseiro, C. Kroer, and R. Kumar, “Online resource allocation under horizon uncertainty,” *arXiv preprint arXiv:2206.13606*, 2022.
- [91] Capital One Shopping Research, *Amazon logistics statistics (2024): Number of package deliveries*, Oct. 2024.
- [92] R. E. Marsten, “An algorithm for large set partitioning problems,” *Management Science*, vol. 20, no. 5, pp. 774–787, 1974.
- [93] M. Ehrgott, *Multicriteria optimization*. Springer Science & Business Media, 2005, vol. 491.
- [94] S. Balseiro, H. Lu, and V. Mirrokni, “Dual mirror descent for online allocation problems,” in *International Conference on Machine Learning*, PMLR, 2020, pp. 613–628.
- [95] G. Mavrotas, “Effective implementation of the  $\epsilon$ -constraint method in multi-objective mathematical programming problems,” *Applied Mathematics and Computation*, vol. 213, no. 2, pp. 455–465, 2009.
- [96] O. Besbes, Y. Kanoria, and A. Kumar, “The fault in our recommendations: On the perils of optimizing the measurable,” in *Proceedings of the 18th ACM Conference on Recommender Systems*, 2024, pp. 200–208.

- [97] A. Klimashevskaja, D. Jannach, M. Elahi, and C. Trattner, “A survey on popularity bias in recommender systems,” *arXiv preprint arXiv:2308.01118*, 2023.
- [98] T. Cunningham *et al.*, “What we know about using non-engagement signals in content ranking,” *arXiv preprint arXiv:2402.06831*, 2024.
- [99] M. Chen *et al.*, “Values of user exploration in recommender systems,” in *Proceedings of the 15th ACM Conference on Recommender Systems*, 2021, pp. 85–95.
- [100] J. Stray, I. Vendrov, J. Nixon, S. Adler, and D. Hadfield-Menell, “What are you optimizing for? aligning recommender systems with human values,” *arXiv preprint arXiv:2107.10939*, 2021.
- [101] B. C. Arnold, “Pareto and generalized pareto distributions,” in *Modeling income distributions and Lorenz curves*, Springer, 2008, pp. 119–145.
- [102] A. B. Ahanger, S. W. Aalam, M. R. Bhat, and A. Assad, “Popularity bias in recommender systems-a review,” in *International Conference on Emerging Technologies in Computer Engineering*, Springer, 2022, pp. 431–444.
- [103] Y.-J. Park and A. Tuzhilin, “The long tail of recommender systems and how to leverage it,” in *Proceedings of the 2008 ACM conference on Recommender systems*, 2008, pp. 11–18.
- [104] Ò. Celma and P. Cano, “From hits to niches? or how popular artists can bias music recommendation and discovery,” in *Proceedings of the 2nd KDD workshop on large-scale recommender systems and the netflix prize competition*, 2008, pp. 1–8.
- [105] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.
- [106] H. Abdollahpouri, R. Burke, and B. Mobasher, “Controlling popularity bias in learning-to-rank recommendation,” in *Proceedings of the eleventh ACM conference on recommender systems*, 2017, pp. 42–46.
- [107] M. D. Ekstrand, A. Das, R. Burke, F. Diaz, *et al.*, “Fairness in information access systems,” *Foundations and Trends® in Information Retrieval*, vol. 16, no. 1-2, pp. 1–177, 2022.
- [108] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, “Bias and debias in recommender system: A survey and future directions,” *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1–39, 2023.
- [109] O. Besbes, Y. Gur, and A. Zeevi, “Optimization in online content recommendation services: Beyond click-through-rates,” *Manufacturing & Service Operations Management*, vol. 18, no. 1, pp. 15–33, 2016.

- [110] T. M. McDonald, L. Maystre, M. Lalmas, D. Russo, and K. Ciosek, “Impatient bandits: Optimizing recommendations for the long-term without delay,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 1687–1697.
- [111] J. Kleinberg, S. Mullainathan, and M. Raghavan, “The challenge of understanding what users want: Inconsistent preferences and engagement optimization,” *Management Science*, 2023.
- [112] S. Milli, E. Pierson, and N. Garg, “Choosing the right weights: Balancing value, strategy, and noise in recommender systems,” *arXiv preprint arXiv:2305.17428*, 2023.
- [113] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.
- [114] A. Anderson, L. Maystre, I. Anderson, R. Mehrotra, and M. Lalmas, “Algorithmic effects on the diversity of consumption on spotify,” in *Proceedings of the web conference 2020*, 2020, pp. 2155–2165.
- [115] J. Sá, V. Queiroz Marinho, A. R. Magalhães, T. Lacerda, and D. Goncalves, “Diversity vs relevance: A practical multi-objective study in luxury fashion recommendations,” in *Proceedings of the 45th International ACM SIGIR Conference on research and development in information retrieval*, 2022, pp. 2405–2409.
- [116] H. Steck, “Calibrated recommendations,” in *Proceedings of the 12th ACM conference on recommender systems*, 2018, pp. 154–162.
- [117] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, “Improving recommendation lists through topic diversification,” in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 22–32.
- [118] P. Castells, N. Hurley, and S. Vargas, “Novelty and diversity in recommender systems,” in *Recommender systems handbook*, Springer, 2021, pp. 603–646.
- [119] H. Yin, B. Cui, J. Li, J. Yao, and C. Chen, “Challenging the long tail recommendation,” *arXiv preprint arXiv:1205.6700*, 2012.
- [120] S. P. Anderson, A. De Palma, and J.-F. Thisse, *Discrete choice theory of product differentiation*. MIT press, 1992.
- [121] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.
- [122] O. Besbes, Y. Kanoria, and A. Kumar, “Impact of rankings and personalized recommendations in marketplaces,” *arXiv preprint arXiv:2506.03369*, 2025.

- [123] Gallup, *On second thought: U.s. adults reflect on their education decisions*, Accessed: 2024-07-31, 2017.
- [124] U.S. News & World Report, *Best colleges rankings*, <https://www.usnews.com/best-colleges>, Accessed: 2024-10-03, 2024.
- [125] National Institutional Ranking Framework (NIRF), *National institutional ranking framework (nirf), ministry of education, government of india*, <https://www.nirfindia.org/>, Accessed: 2024-10-03, 2024.
- [126] J. Chen *et al.*, “When large language models meet personalization: Perspectives of challenges and opportunities,” *World Wide Web*, vol. 27, no. 4, p. 42, 2024.
- [127] Z. Zhang *et al.*, “Personalization of large language models: A survey,” *arXiv preprint arXiv:2411.00027*, 2024.
- [128] A. Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.
- [129] J. B. Schafer, J. Konstan, and J. Riedl, “Recommender systems in e-commerce,” in *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 158–166.
- [130] G. Häubl and V. Trifts, “Consumer decision making in online shopping environments: The effects of interactive decision aids,” *Marketing science*, vol. 19, no. 1, pp. 4–21, 2000.
- [131] G. Adomavicius, Z. Huang, and A. Tuzhilin, “Personalization and recommender systems,” in *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, INFORMS, 2008, pp. 55–107.
- [132] S. Berkovsky, T. Kuflik, and F. Ricci, “Mediation of user models for enhanced personalization in recommender systems,” *User Modeling and User-Adapted Interaction*, vol. 18, pp. 245–286, 2008.
- [133] M. Naumov *et al.*, “Deep learning recommendation model for personalization and recommendation systems,” *arXiv preprint arXiv:1906.00091*, 2019.
- [134] Y. Su, M. Bayoumi, and T. Joachims, “Optimizing rankings for recommendation in matching markets,” in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 328–338.
- [135] A. Aouad and D. Saban, “Online assortment optimization for two-sided matching platforms,” *Management Science*, vol. 69, no. 4, pp. 2069–2087, 2023.
- [136] P. Shi, “Optimal match recommendations in two-sided marketplaces with endogenous prices,” *Management Science*, 2024.

- [137] D. Gale and L. S. Shapley, “College admissions and the stability of marriage,” *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
- [138] A. E. Roth and M. Sotomayor, “Two-sided matching,” *Handbook of game theory with economic applications*, vol. 1, pp. 485–541, 1992.
- [139] A. Abdulkadiroğlu and T. Sönmez, “School choice: A mechanism design approach,” *American economic review*, vol. 93, no. 3, pp. 729–747, 2003.
- [140] S. Campbell, L. Macmillan, R. Murphy, and G. Wyness, “Matching in the dark? inequalities in student to degree match,” *Journal of Labor Economics*, vol. 40, no. 4, pp. 807–850, 2022.
- [141] E. W. Dillon and J. A. Smith, “Determinants of the match between student ability and college quality,” *Journal of Labor Economics*, vol. 35, no. 1, pp. 45–66, 2017.
- [142] N. Immorlica, J. Leshno, I. Lo, and B. Lucier, “Information acquisition in matching markets: The role of price discovery,” *Available at SSRN 3705049*, 2020.
- [143] Y. Chen and Y. He, “Information acquisition and provision in school choice: An experimental study,” *Journal of Economic Theory*, vol. 197, p. 105 345, 2021.
- [144] J. Grenet, Y. He, and D. Kübler, “Preference discovery in university admissions: The case for dynamic multioffer mechanisms,” *Journal of Political Economy*, vol. 130, no. 6, pp. 1427–1476, 2022.
- [145] S. P. Corcoran, J. L. Jennings, S. R. Cohodes, and C. Sattin-Bajaj, “Leveling the playing field for high school choice: Results from a field experiment of informational interventions,” National Bureau of Economic Research, Tech. Rep., 2018.
- [146] S. R. Cohodes, S. P. Corcoran, J. L. Jennings, and C. Sattin-Bajaj, “When do informational interventions work? experimental evidence from new york city high school choice,” *Educational Evaluation and Policy Analysis*, vol. 47, no. 1, pp. 208–236, 2025.
- [147] C. M. Hoxby and S. Turner, “What high-achieving low-income students know about college,” *American Economic Review*, vol. 105, no. 5, pp. 514–517, 2015.
- [148] T. Larroucau, I. Rios, A. Fabre, and C. Neilson, “College application mistakes and the design of information policies at scale,” 2024.
- [149] M. Elliott, A. Galeotti, A. Koh, and W. Li, “Matching and information design in marketplaces,” *Available at SSRN 4283968*, 2022.
- [150] K. Bimpikis, Y. Papanastasiou, and W. Zhang, “Information provision in two-sided platforms: Optimizing for supply,” *Management Science*, vol. 70, no. 7, pp. 4533–4547, 2024.

- [151] Y. Papanastasiou, K. Bimpikis, and N. Savva, “Crowdsourcing exploration,” *Management Science*, vol. 64, no. 4, pp. 1727–1746, 2018.
- [152] S. Dasgupta, “Designing information to improve welfare in matching markets,” *Mathematical Social Sciences*, 2024.
- [153] J. Kleinberg and M. Raghavan, “Algorithmic monoculture and social welfare,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 22, e2018340118, 2021.
- [154] K. Peng and N. Garg, “Monoculture in matching markets,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 81 959–81 991, 2024.
- [155] K. Peng and N. Garg, “Wisdom and foolishness of noisy matching markets,” *arXiv preprint arXiv:2402.16771*, 2024.
- [156] J. Baek, H. Bastani, and S. Chen, “Hiring under congestion and algorithmic monoculture: Value of strategic behavior,” *arXiv preprint arXiv:2502.20063*, 2025.
- [157] S. Baswana, P. P. Chakrabarti, S. Chandran, Y. Kanoria, and U. Patange, “Centralized admissions for engineering colleges in india,” in *Proceedings of the 2019 ACM Conference on Economics and Computation*, 2019, pp. 323–324.
- [158] A. Abdulkadiroğlu and T. Sönmez, “Random serial dictatorship and the core from random endowments in house allocation problems,” *Econometrica*, vol. 66, no. 3, pp. 689–701, 1998.
- [159] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.
- [160] J. Nair, A. Wierman, and B. Zwart, *The fundamentals of heavy tails: Properties, emergence, and estimation*. Cambridge University Press, 2022, vol. 53.
- [161] G. Monge, “Mémoire sur la théorie des déblais et des remblais,” *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.
- [162] S. Hundrieser, T. Staudt, and A. Munk, “Empirical optimal transport between different measures adapts to lower complexity,” *arXiv preprint arXiv:2202.10434*, 2022.
- [163] S. Jasin and S. Kumar, “A re-solving heuristic with bounded revenue loss for network revenue management with customer choice,” *Mathematics of Operations Research*, vol. 37, no. 2, pp. 313–345, 2012.
- [164] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2019, vol. 49.
- [165] A. E. Taylor, *General theory of functions and integration*. Courier Corporation, 1985.

## Appendix A: Feature-Based Dynamic Matching

### A.1 Proof of $U_\infty(P, Q, \varphi) \geq U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$

That  $U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$  is straightforward. To prove the result it suffices to show the first inequality, i.e.  $U_\infty(P, Q, \varphi) \geq U_n^H(P, Q, \varphi)$ , which follows from the next Lemma. With bounded quality function  $\varphi$ ,  $U_n^H$  are trivially bounded, then by monotone convergence theorem  $U_\infty(P, Q, \varphi)$  is also bounded, making it a valid performance benchmark.

**Lemma 5**  $U_n^H, n \geq 1$  is a monotone increasing sequence.

*Proof of Lemma 5.* For each realized  $(X_t, Y_{(t)})_{1 \leq t \leq n}$ , we consider a specific randomized permutation. Simulate  $n - 1$  demand units  $X'_1, \dots, X'_{n-1}$  *i.i.d.* from  $P$ . Let  $\lambda$  denote the optimal assignment between  $\{X'_1, \dots, X'_{n-1}, X_n\}$  and  $\{Y_1, \dots, Y_n\}$ . We then optimally match  $\{X_1, \dots, X_{n-1}\}$  and  $\{Y_1, \dots, Y_n\} / \{Y_{\lambda_n}\}$ . The above constitutes a permutation (possibly randomized) between  $\{X_1, \dots, X_n\}$  and  $\{Y_1, \dots, Y_n\}$ , which we call  $\sigma'$ . Notice that

$$\begin{aligned} U_n^H &= \frac{1}{n} \mathbb{E} \left[ \sup_{\sigma \in S_n} \sum_{t=1}^n \varphi(X_t, Y_{\sigma_t}) \right] \geq \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \varphi(X_t, Y_{\sigma'_t}) \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^{n-1} \varphi(X_t, Y_{\sigma'_t}) \right] + \frac{1}{n} \mathbb{E} [\varphi(X_n, Y_{\sigma'_n})] \\ &= \frac{n-1}{n} U_{n-1}^H + \frac{1}{n} U_n^H, \end{aligned}$$

where the last step follows from (a)  $X_1, \dots, X_{n-1}$  is *i.i.d.*  $P$ , (b)  $\{Y_1, \dots, Y_n\} / \{Y_{\lambda_n}\}$  is *i.i.d.*  $Q$  and (c)  $(X_n, Y_{\lambda_n})$  is a pair taken from the optimal offline assignment of size  $n$ . The desired result thus follows. ■

**Remark 8** We provide an example to highlight the distinction between the limiting hindsight optimum and the fluid optimum, which in our setting is the optimal transport value between the

distributions  $P$  and  $Q$  w.r.t. function  $\varphi$ . Suppose the supply and demand distributions  $P = Q = \text{Uniform}([0, 1])$ , and the quality function is  $\varphi(X, Y) = \mathbb{1}\{X = Y\}$ , then for any finite number of supply and demand units  $n$ , the hindsight optimum value  $U_n(P, Q, \varphi) = 0$ . Consequently,  $U_\infty(P, Q, \varphi) = 0$  whereas the fluid optimum is 1.

## A.2 Proof of the Failure of Greedy in Section 1.3.1

### A.2.1 Proof of Proposition 1

*Proof of Proposition 1.* We assume that the quality function is  $\varphi(X, Y) = -|X - Y|^p$ . Instead of taking the quality function maximizing perspective, we will take the cost minimization perspective and consider the cost function  $d(X, Y) = |X - Y|^p$ . Note that we will use  $U_n(\text{Greedy}; P, Q)$ ,  $U_n^H(P, Q)$  and  $U_\infty(P, Q)$  to denote the average matching cost under Greedy, the hindsight optimal cost and the limiting hindsight cost. Note that  $U_n^H(P, Q) \geq U_\infty(P, Q)$  and hence we have that

$$\text{REG}_n(\text{Greedy}; P, Q) = U_n(\text{Greedy}; P, Q) - U_\infty(P, Q) \geq U_n(\text{Greedy}; P, Q) - U_n^H(P, Q) \quad (\text{A.1})$$

Note that since  $f_p(x) > 0$  for all  $x \in [0, 1]$ , we have that  $F_P$  and  $F_P^{-1}$  are strictly increasing functions over the interval  $[0, 1]$ , where  $F_P^{-1}(y) = \{x : F_P(x) = y\}$ . Fix  $\rho \in (0, 1)$  and let  $F_P(0.5) := \alpha \neq 1 - \rho$  and let  $\bar{\alpha} := \bar{F}_P(0.5) = 1 - F_P(0.5)$ . Let us assume that  $\alpha < 1 - \rho$ . Note that this assumption is without any loss because if  $\alpha > 1 - \rho$ , then the entire analysis can be symmetrically made for the supply units located at 0.

Define  $r \triangleq \rho / (1 - \alpha)$ . Note that since  $\alpha < 1 - \rho$  (by assumption), we have that  $r < 1$ . Define  $\delta \triangleq 1 - \alpha - \rho > 0$ . Note that  $(1 - r)\rho = r\delta$ . Define  $\underline{x}^* \triangleq F_P^{-1}(\alpha + \delta/2)$ ,  $x^* \triangleq F_P^{-1}(1 - \rho)$  and  $\bar{x}^* \triangleq F_P^{-1}(1 - \rho/2)$ . Since  $\alpha < \alpha + \delta/2 < 1 - \rho < 1 - \rho/2$  and  $F_P^{-1}$  is strictly increasing, we have that  $0.5 < \underline{x}^* < x^* < \bar{x}^*$ .

For  $x_2 > x_1$  and  $t_2 > t_1$ , define  $N_{t_1:t_2}^X(x_1, x_2) \triangleq \sum_{k=t_1}^{t_2} \mathbb{1}\{X_k \in (a, b)\}$  to be the number of demand

units that arrive in the time interval  $\{t_1, t_1 + 1, \dots, t_2\}$  and are located in the interval  $(x_1, x_2)$ . Let  $N_1^Y = \sum_{k=1}^n \mathbb{1}\{Y_k = 1\}$  denote the number of supply units located at 1.

We will now define some events

$$\begin{aligned}
E^Y &= \{\rho n - \sqrt{n}/16 \leq N_1^Y \leq \rho n - \sqrt{n}/8\} \\
L_1^X &= \{\delta r n/2 - \sqrt{n}/32 \leq N_{1:[rn]}^X(0.5, \underline{x}^*) \leq \delta r n/2 + \sqrt{n}/32\} \\
L_2^X &= \{\delta r n/2 - \sqrt{n}/32 \leq N_{1:[rn]}^X(\underline{x}^*, x^*) \leq \delta r n/2 + \sqrt{n}/32\} \\
L_3^X &= \{\rho r n/2 - \sqrt{n}/32 \leq N_{1:[rn]}^X(x^*, \bar{x}^*) \leq \rho r n/2 + \sqrt{n}/32\} \\
L_4^X &= \{\rho r n/2 - \sqrt{n}/32 \leq N_{1:[rn]}^X(\bar{x}^*, 1) \leq \rho r n/2 + \sqrt{n}/32\} \\
H_1^X &= \{\delta(1-r)n/2 - \sqrt{n}/32 \leq N_{[rn]+1:n}^X(0.5, \underline{x}^*) \leq \delta(1-r)n/2 + \sqrt{n}/32\} \\
H_2^X &= \{\delta(1-r)n/2 - \sqrt{n}/32 \leq N_{[rn]+1:n}^X(\underline{x}^*, x^*) \leq \delta(1-r)n/2 + \sqrt{n}/32\} \\
H_3^X &= \{\rho(1-r)n/2 - \sqrt{n}/32 \leq N_{[rn]+1:n}^X(x^*, \bar{x}^*) \leq \rho(1-r)n/2 + \sqrt{n}/32\} \\
H_4^X &= \{\rho(1-r)n/2 - \sqrt{n}/32 \leq N_{[rn]+1:n}^X(\bar{x}^*, 1) \leq \rho(1-r)n/2 + \sqrt{n}/32\}
\end{aligned}$$

We have that  $N_1^Y \sim \text{Binomial}(n, \rho)$  and hence using CLT there exists a constant  $c > 0$  such that  $\mathbb{P}(E^Y) \geq c$ . We have that  $\mathbb{P}(X \in (0.5, \underline{x}^*)) = F_P(\underline{x}^*) - F_P(0.5) = \delta/2$ . Similarly, we have that  $\mathbb{P}(X \in (\underline{x}^*, x^*)) = \delta/2$ ,  $\mathbb{P}(X \in (x^*, \bar{x}^*)) = \rho/2$  and  $\mathbb{P}(X \in (\bar{x}^*, 1)) = \rho/2$ . Therefore, using standard CLT arguments, we have that there exists a constant  $c > 0$  such that  $\mathbb{P}(L_i^X) \geq c$  for all  $i \in \{1, 2, 3, 4\}$  and  $\mathbb{P}(H_i^X) \geq c$  for all  $i \in \{1, 2, 3, 4\}$ . Define  $G = E^Y \cap (\cap_{i=1}^4 L_i^X) \cap (\cap_{i=1}^4 H_i^X)$ , then using standard conditioning arguments (refer to [43]), we have that for some constant  $\beta > 0$ ,  $\mathbb{P}(G) \geq \beta$ .

**Hindsight Optimum.** Under the event  $G$ , we have that  $N_{1:n}^X(x^*, 1) \geq \rho n - \sqrt{n}/8$  and  $N_1^Y \leq \rho n - \sqrt{n}/8$ . Therefore, under the event  $G$ , the hindsight optimal will match all the supply units located to 1 to the demand units that arrive in the interval  $(x^*, 1)$ . Note that since  $N_{1:n}^X(x^*, 1) \geq \rho n - \sqrt{n}/8$ , not all demand units in the interval  $(x^*, 1)$  will be matched to the supply units located at 1. The demand units not matched to the supply units located 1 will be matched to the supply units located

at 0.

**Greedy Algorithm.** Under the event  $G$ , we have that  $N_{1:[rn]}^X(0.5, 1) \geq \rho n - \sqrt{n}/8$  and  $N_1^Y \leq \rho n - \sqrt{n}/8$ . Therefore under the event  $G$ , the Greedy algorithm will match all the supply units located at 1 to the demand units that arrive in the interval  $(0.5, 1)$  up till the time  $t = \lfloor rn \rfloor$ . For  $t \geq \lfloor rn \rfloor + 1$ , Greedy will match all the demand units to the supply units located at 0. This includes the demand units that will arrive in the interval  $(x^*, 1)$ . The two places that Greedy differs from the hindsight optimal matching is:

- (i) at least  $\delta rn - \sqrt{n}/8$  of the demand arrivals in the interval  $(0.5, x^*)$  during the first  $\lfloor rn \rfloor$  time steps are matched to the supply units at 1 under Greedy whereas they are matched to the supply units at 0 under the hindsight optimal matching. This is because all the arrivals in  $(0.5, x^*)$  are matched under the hindsight optimal to the supply units at 0. Under Greedy, we have that at most  $\rho n + \sqrt{n}/16$  arrivals in the interval  $(x^*, 1)$  get matched to supply units at 1 and since there are at least  $\rho n - \sqrt{n}/16$  supply units at 1, we have the at least  $\delta rn - \sqrt{n}/8$  of the arrivals in  $(0.5, x^*)$  get matched to supply units at 1.
- (ii) at least  $\rho(1-r)n - \sqrt{n}/8$  of the demand arrivals in the interval  $(x^*, 1)$  during the last  $n - \lfloor rn \rfloor$  time steps are matched to the supply units at 0 under Greedy whereas under the hindsight optimal matches these demand units are matched the supply units at 1. This is because all the demand arrivals in the last  $n - \lfloor rn \rfloor$  time steps are matched the supply units at 0 under Greedy since Greedy prematurely exhausts all the supply units at 1. Since there are  $\rho n - \sqrt{n}/16$  supply units at location 1 and at most  $\rho n + \sqrt{n}/16$  of the demand arrivals in the interval  $(x^*, 1)$  get matched in first  $\lfloor rn \rfloor$  time steps, we have that at least  $\rho(1-r)n - \sqrt{n}/8$  demand units are available to be matched in the last  $n - \lfloor rn \rfloor$  time steps.

Let  $\pi^g$  denote the allocation under the Greedy algorithm. Then we have that

$$\begin{aligned}
\text{REG}_n(\text{Greedy}; P, Q) &\stackrel{(a)}{\geq} U_n(\text{Greedy}; P, Q) - U_n^H(P, Q), \\
&\stackrel{(b)}{=} n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \right], \\
&\stackrel{(c)}{=} n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \middle| G \right] \mathbb{P}(G) \\
&\quad + n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \middle| G^c \right] \mathbb{P}(G^c), \\
&\stackrel{(d)}{\geq} \beta n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \middle| G \right],
\end{aligned}$$

where (a) follows from (B.14), (b) from the definition of  $U_n(\text{Greedy}; P, Q)$  and  $U_n^H(P, Q)$ , (c) follows from law of total expectation and (d) follows from the fact that  $\sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \geq 0$ ,  $\mathbb{P}(G^c) \geq 0$  and  $\mathbb{P}(G) \geq \beta$ . Now it suffices to show that there exists a constant  $\kappa > 0$  such that

$$n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \middle| G \right] \geq \kappa$$

Given the supply and demand units  $\{Y_1, Y_2, \dots, Y_n\}$  and  $\{X_1, X_2, \dots, X_n\}$ , let  $\sigma(X)$  and  $\pi^g(X)$  denote the hindsight optimal assignment and the Greedy assignment of demand unit  $X$  respectively. Furthermore, let  $\mathcal{A}_{t_1:t_2}(a, b) = \{X_k : k \in \{t_1, t_1 + 1, \dots, t_2\} \text{ and } X_k \in (a, b)\}$ . Now we have can re-write the summation  $\sum_{t=1}^n c(X_t, Y_{\pi_t^g})$  and  $\sum_{t=1}^n c(X_t, Y_{\sigma_t})$  as follows

$$\begin{aligned}
\sum_{t=1}^n c(X_t, Y_{\pi_t^g}) &= \sum_{X \in \mathcal{A}_{1:n}(0,0.5)} c(X, \pi^g(X)) + \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \pi^g(X)) \\
\sum_{t=1}^n c(X_t, Y_{\sigma_t}) &= \sum_{X \in \mathcal{A}_{1:n}(0,0.5)} c(X, \sigma(X)) + \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \sigma(X))
\end{aligned}$$

Note that under the event  $G$ , we have that  $c(X, \pi^g(X)) = c(X, \sigma(X))$  for all  $X \in (0, 0.5)$  and therefore we have that under the event  $G$ ,

$$\sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) = \sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} c(X, \pi^g(X)) - \sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} c(X, \sigma(X))$$

For the Greedy algorithm we have that,

$$\begin{aligned} \sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} c(X, \pi^g(X)) &= \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5, x^*)} c(X, \pi^g(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(0.5, x^*)} c(X, \pi^g(X)) \\ &+ \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(x^*, 1)} c(X, \pi^g(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, 1)} c(X, \pi^g(X)) \end{aligned}$$

For the hindsight optimal we have that,

$$\begin{aligned} \sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} c(X, \sigma(X)) &= \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5, x^*)} c(X, \sigma(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(0.5, x^*)} c(X, \sigma(X)) \\ &+ \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(x^*, 1)} c(X, \sigma(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, 1)} c(X, \sigma(X)) \end{aligned}$$

Under the event  $G$ , we have that for all  $X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(0.5, x^*)$ , we have that  $c(X, \pi^g(X)) = c(X, \sigma(X))$  and we have that  $|\sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(x^*, 1)} c(X, \pi^g(X)) - c(X, \sigma(X))| \leq \sqrt{n}$ .

Therefore we have that

$$\begin{aligned} \sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} (c(X, \pi^g(X)) - c(X, \sigma(X))) &\geq \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5, x^*)} c(X, \pi^g(X)) - c(X, \sigma(X)) \\ &+ \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, 1)} c(X, \pi^g(X)) - c(X, \sigma(X)) - \sqrt{n} \end{aligned}$$

Next we will provide a lower bound for the sum  $\sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5, x^*)} (c(X, \pi^g(X)) - c(X, \sigma(X)))$ .

Consider the demand arrivals in  $\mathcal{A}_{1:\lfloor rn \rfloor}(0.5, \underline{x}^*)$ , we have that  $c(X, \pi^g(X)) \geq (1 - \underline{x}^*)^p$  and  $c(X, \sigma(X)) \leq (\underline{x}^*)^p$  and this is true for at least  $\delta rn/2 - \sqrt{n}$  arrivals. Similarly, consider the demand arrivals in  $\mathcal{A}_{1:\lfloor rn \rfloor}(\underline{x}^*, x^*)$ , we have that  $c(X, \pi^g(X)) \geq (1 - x^*)^p$  and  $c(X, \sigma(X)) \leq (x^*)^p$  and

this is also true for at least  $\delta rn/2 - \sqrt{n}$ . Therefore we have that

$$\sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5, x^*)} c(X, \pi^g(X)) - c(X, \sigma(X)) \geq [(1 - \underline{x}^*)^p + (1 - x^*)^p - (\underline{x}^*)^p - (x^*)^p](\delta rn/2) - 2\sqrt{n}$$

Next we will provide a lower bound for the sum  $\sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, 1)} (c(X, \pi^g(X)) - c(X, \sigma(X)))$ .

Consider the demand arrivals in  $\mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, \bar{x}^*)$ , we have that  $c(X, \pi^g(X)) \geq (x^*)^p$  and  $c(X, \sigma(X)) \leq (1 - x^*)^p$  and this is true for at least  $\rho(1 - r)n/2 - \sqrt{n}$  arrivals. Similarly, consider the demand arrivals in  $\mathcal{A}_{\lfloor rn \rfloor+1:n}(\bar{x}^*, 1)$ , we have that  $c(X, \pi^g(X)) \geq (\bar{x}^*)^p$  and  $c(X, \sigma(X)) \leq (1 - \bar{x}^*)^p$  and this is true for at least  $\rho(1 - r)n/2 - \sqrt{n}$  arrivals. Therefore we have that

$$\begin{aligned} \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^*, 1)} (c(X, \pi^g(X)) - c(X, \sigma(X))) &\geq [-(1 - \bar{x}^*)^p - (1 - x^*)^p + (\bar{x}^*)^p + (x^*)^p](\rho(1 - r)n/2) \\ &\quad - 2\sqrt{n} \end{aligned}$$

Define  $\beta' = (1 - \underline{x}^*)^p - (1 - \bar{x}^*)^p + (\bar{x}^*)^p - (\underline{x}^*)^p > 0$ . Since  $\rho(1 - r) = \delta r$ , we have that

$$\sum_{X \in \mathcal{A}_{1:n}(0.5, 1)} (c(X, \pi^g(X)) - c(X, \sigma(X))) \geq \beta' \cdot \frac{\rho(1 - \rho - \alpha)}{1 - \alpha} \cdot n - 5\sqrt{n}$$

Therefore we have that

$$n^{-1} \mathbb{E} \left[ \sum_{t=1}^n c(X_t, Y_{\pi_t^g}) - \min_{\sigma} \sum_{t=1}^n c(X_t, Y_{\sigma_t}) \middle| G \right] \geq \beta' \frac{\rho}{1 - \alpha} (1 - \rho - \alpha) - 5/\sqrt{n}$$

This concludes the proof as we can choose a large enough  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ , the lower bound in the Proposition 1 holds. ■

### A.3 Proof of Corollary 1

*Proof of Corollary 1.* From the proof of Lemma 5,  $U_\infty(P, Q, \varphi) = \lim_{n \rightarrow \infty} U_n^H(P, Q, \varphi)$  exists.

Equivalently,  $\{\text{REG}_k(\text{H-OPT})\}_{k \geq 1}$  is a non-negative monotone decreasing sequence and

$$\lim_{k \rightarrow \infty} \text{REG}_k(\text{H-OPT}) = 0.$$

By Remark 1, we have  $\text{REG}_n(\text{SOAR}) = \frac{1}{n} \sum_{k=1}^n \text{REG}_k(\text{H-OPT})$ , which also converges to 0 as  $n \rightarrow \infty$ . ■

#### A.4 Proof of Corollary 2

*Proof of Corollary 2.* First, we show  $\beta \leq 1$ . By definition,  $\text{REG}_n(\text{H-OPT}) = U_\infty - U_n^{\text{H}}$ . As the cumulative regret is at least a constant,  $U_\infty - U_n^{\text{H}} \geq \frac{c}{n}$ . Take lim sup on both sides,  $\limsup_{n \rightarrow \infty} n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\text{H}}) \geq \limsup_{n \rightarrow \infty} n^{\beta-1-\epsilon} = 0$ , thus  $\beta < 1 + \epsilon$  for all  $\epsilon > 0$ .

1)  $\limsup_{n \rightarrow \infty} n^{\beta-\epsilon} \cdot \text{REG}_n(\text{SOAR}) = 0$ .

As  $\limsup_{n \rightarrow \infty} n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\text{H}}) = 0$ , for any  $\delta > 0$ , there exists  $N_\delta \in \mathbb{N}$  such that  $n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\text{H}}) < \delta$  for  $n \geq N_\delta$ . Then we have for all  $n \in \mathbb{N}$ ,

$$\text{REG}_n(\text{SOAR}) \stackrel{(a)}{=} U_\infty - U_n(\text{SOAR}) \stackrel{(b)}{=} \frac{1}{n} \sum_{k=1}^n (U_\infty - U_k^{\text{H}}) < \frac{N_\delta U_\infty}{n} + \frac{\delta}{n} \sum_{k=N_\delta+1}^n k^{-\beta+\epsilon} \stackrel{(c)}{<} \frac{N_\delta U_\infty}{n} + \frac{\delta / (-\beta + \epsilon + 1)}{n^{\beta-\epsilon}},$$

where (a) follows by definition; (b) follows Theorem 1; and (c) follows as  $\epsilon < \beta$ , and  $\sum_{k=N_\delta+1}^n k^{-\beta+\epsilon} \leq \int_{N_\delta}^n k^{-\beta+\epsilon} dk$ . Therefore, for any  $\delta' > 0$ , let  $N_{\delta'}$  be the minimum  $n$  such that  $N_\delta U_\infty n^{\beta-\epsilon-1} + \frac{\delta}{-\beta+\epsilon+1} < \delta'$ , where the existence of  $N_{\delta'}$  follows from  $\beta < 1 + \epsilon$ . Then we have for all  $n \geq N_{\delta'}$ ,

$$0 < n^{\beta-\epsilon} \cdot \text{REG}_n(\text{SOAR}) < N_\delta U_\infty n^{\beta-\epsilon-1} + \frac{\delta}{-\beta + \epsilon + 1} < \delta'.$$

Therefore  $\limsup_{n \rightarrow \infty} n^{\beta-\epsilon} \cdot \text{REG}_n(\text{SOAR}) = 0$ .

2)  $\limsup_{n \rightarrow \infty} n^{\beta+\epsilon} \cdot \text{REG}_n(\text{SOAR}) = \infty$ .

As  $\liminf_{n \rightarrow \infty} n^{\beta+\epsilon} \cdot (U_\infty - U_n^{\text{H}}) = \infty$ , for any  $M > 0$ , there exists  $N_M \in \mathbb{N}$  such that  $n^{\beta+\epsilon} \cdot$

$\text{REG}_n(\text{H-OPT}) > M$  for  $n \geq N_M$ . Then we have for all  $n > N_M$ ,

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &> \frac{1}{n} \sum_{k=1}^{N_M} \frac{C}{k} + \frac{1}{n} \sum_{k=N_M+1}^n \frac{M}{k^{\beta+\epsilon}} > \frac{1}{n} \int_1^{N_M} \frac{C}{k} dk + \frac{1}{n} \int_{N_M+1}^n \frac{M}{k^{\beta+\epsilon}} dk \\ &= \frac{C \log(N_M)}{n} + \frac{M/(1-\beta-\epsilon)}{n} (n^{1-\beta-\epsilon} - (N_M+1)^{1-\beta-\epsilon}). \end{aligned}$$

- If  $\beta + \epsilon > 1$ , then

$$n^{\beta+\epsilon} \cdot \text{REG}_n(\text{SOAR}) > C \log(N_M) n^{\beta+\epsilon-1},$$

thus  $\liminf_{n \rightarrow \infty} n^{\beta+\epsilon} \text{REG}_n(\text{SOAR}) = \infty$ .

- If  $\beta + \epsilon < 1$ , then

$$n^{\beta+\epsilon} \cdot \text{REG}_n(\text{SOAR}) > \frac{M}{1-\beta-\epsilon} (1 - (N_M+1)/n)^{1-\beta-\epsilon}.$$

Therefore, for any  $M' > 0$ , let  $N_{M'}$  be the minimum  $n$  such that  $\frac{M}{1-\beta-\epsilon} (1 - (N_M+1)/n)^{1-\beta-\epsilon} > M'$ , where the existence of  $N_{M'}$  follows from  $\beta + \epsilon < 1$ . Then for all  $n \geq N_{M'}$ ,

$$n^{\beta+\epsilon} \cdot \text{REG}_n(\text{SOAR}) > M'.$$

Therefore  $\limsup_{n \rightarrow \infty} n^{\beta+\epsilon} \cdot \text{REG}_n(\text{SOAR}) = \infty$ .

### 3) $\text{REG}_n(\text{SOAR}) \leq n^\epsilon \cdot \text{REG}_n(\text{H-OPT})$ for $n$ sufficiently large.

We have shown that

$$\text{REG}_n(\text{SOAR}) < \frac{N_\delta U_\infty}{n} + \frac{\delta/(-\beta + \epsilon + 1)}{n^{\beta-\epsilon}} = \Theta(n^{-\beta+\epsilon}) \text{ for } n \in \mathbb{N},$$

and for any  $M > 0$ , there exists  $N_M \in \mathbb{N}$  such that

$$\text{REG}_n(\text{H-OPT}) > M n^{-\beta-\epsilon} \text{ for } n \geq N_M.$$

Combining them gives  $\text{REG}_n(\text{SOAR}) \leq n^\epsilon \cdot \text{REG}_n(\text{H-OPT})$  for  $n$  sufficiently large.

### Polynomial Regret Scaling Case.

If in addition,  $\lim_{n \rightarrow \infty} n^\beta \cdot (U_\infty - U_n^H) = l_0$ , i.e. for any  $\delta > 0$ , there exists  $N$  such that for any  $n > N$ ,  $|n^\beta \cdot \text{REG}_n(\text{H-OPT}) - l_0| < \delta$ . Applying Theorem 1, we have for all  $n \in \mathbb{N}$ ,

$$\text{REG}_n(\text{SOAR}) \leq \frac{NU_\infty}{n} + \frac{1}{n} \sum_{k=N}^n \frac{l_0 + \delta}{k^\beta} = \begin{cases} \Theta(n^{-\beta}), & \beta \neq 1, \\ \Theta(n^{-1} \log n), & \beta = 1. \end{cases}$$

Therefore we have

$$\text{REG}_n(\text{SOAR}) \leq \begin{cases} l_1 \text{REG}_n(\text{H-OPT}), & \beta \neq 1, \\ l_1 \log n \cdot \text{REG}_n(\text{H-OPT}), & \beta = 1. \end{cases}$$

Similarly we have for  $n$  sufficiently large,

$$\text{REG}_n(\text{SOAR}) \geq \frac{1}{n} \sum_{k=1}^N \frac{C}{k} + \frac{1}{n} \sum_{k=N+1}^n \frac{l_0 - \delta}{k^\beta} \begin{cases} \Theta(n^{-\beta}), & \beta \neq 1, \\ \Theta(n^{-1} \log n), & \beta = 1. \end{cases}$$

Combining the above gives the tight characterization of the scaling of  $\text{REG}_n(\text{SOAR})$ . ■

## A.5 Examples of Matching Instances Scale Regularly

In this section, we list some problem settings when the offline optima scale regularly. First, we introduce two useful results.

**Theorem 13 (Eq. (6) and (25) of [52])** *Suppose a set of  $n$  demand points and  $n$  supply points are generated independently and uniformly at random in the hypercube  $[0, 1]^d$ , with matching cost*

$\|X - Y\|^p$ . Then the average cost of the optimal assignment, denoted by  $U_n^H(p, d)$ , is given by

$$U_n^H(2, d) \approx \begin{cases} \frac{1}{6n} + \frac{e_1^{(2)}}{n^2}, & d = 1, \\ \frac{1}{2\pi} \frac{\ln n}{n} + \frac{e_2^{(2)}}{n}, & d = 2, \\ e_d^{(2)} n^{-\frac{2}{d}} + \frac{\zeta_d(1)}{2\pi^2} n^{-1}, & d > 2. \end{cases}$$

Here and in the following the symbol  $\approx$  means that the term on the l.h.s. is asymptotically equal to the r.h.s. except for some additional terms decaying faster than each term in the r.h.s. (e.g.  $U_n^H(2, 1) = \frac{1}{6n} + \frac{e_1^{(2)}}{n^2} + o\left(\frac{1}{n^2}\right)$ ). Conjectured based on numerical simulations:

$$U_n^H(p, d) \approx e_d^{(p)} n^{-\frac{p}{d}} + \alpha_d^{(p)} n^{\frac{2-p-d}{d}} \quad \text{for } d > 2, p > 0,$$

where coefficients  $e_d^{(p)}, \alpha_d^{(p)}$  are constants and  $\zeta_d(x)$  is the Epstein zeta function.

By Definition 1, as  $U_\infty = 0$ , and  $U_n^H(p, d) = \tilde{\Theta}(n^{-p/d})$ , it is easy to see this matching instance scale regularly with  $\beta = p/d$ . For example,  $\lim_{n \rightarrow \infty} n \cdot U_n^H(2, 1) = \Theta(1)$ , thus SOAR gives a tight regret scaling.

## A.6 Details Related to Optimal Transport and Useful Known Results

In this section, we will provide some additional notation and some existing results in the empirical optimal transport literature which will leverage to prove the theorems in Section 1.4.

### A.6.1 Background on Optimal Transport and Wasserstein- $p$ distance

In this section, we provide some background on optimal transport and Wasserstein- $p$  distance. Some of the notations are adapted from [60, Section 2]. For a fixed  $d \geq 1$ , let  $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$  and let  $\mathcal{P}(\mathcal{X})$  denote the set of Borel probability measures with support contained in  $\mathcal{X}$ . Let  $P \in \mathcal{P}(\mathcal{X})$  and  $Q \in \mathcal{P}(\mathcal{Y})$ . An optimal transport map  $\mathcal{T}_{P \rightarrow Q}$  from distribution  $P$  to  $Q$  (with support in the set

$\Omega$ ) is any solution to the *Monge problem* [161] defined as

$$\arg \min_{\mathcal{T}_{P \rightarrow Q} \in \mathcal{T}(P, Q)} \int_{\Omega} \|x - \mathcal{T}_{P \rightarrow Q}(x)\|^2 dP(x), \quad (\text{A.2})$$

where  $\mathcal{T}(P, Q)$  is the set of all transport maps between  $P$  and  $Q$ , i.e., the set of Borel-measurable functions  $\mathcal{T} : \Omega \rightarrow \Omega$  such that  $\mathcal{T}_{\#}P = Q$ . Here,  $\mathcal{T}_{\#}Q$  denotes the pushforward measure of  $Q$  induced by  $\mathcal{T}$ . The convex relaxation of the *Monge problem* is the *Kantorovich problem*,

$$\arg \min_{\pi \in \Pi(P, Q)} \int_{\Omega} \|x - y\|^2 d\pi(x, y), \quad (\text{A.3})$$

where  $\Pi(P, Q)$  is the set of couplings of probability distributions  $P$  and  $Q$ , i.e., the set of probability distributions over  $\mathcal{X} \times \mathcal{Y}$  with first and second margins being  $P$  and  $Q$ . Therefore, a probability measure  $\pi$  over  $\mathcal{X} \times \mathcal{Y}$  belongs to  $\Pi(P, Q)$  if and only if  $\pi(A \times \mathcal{Y}) = P(A)$  and  $\pi(\mathcal{X} \times B) = Q(B)$  holds for every subset  $A$  of  $\mathcal{X}$  and  $B$  of  $\mathcal{Y}$ . It can be shown that a minimizer  $\pi$  always exists for (A.3) [64, Theorem 4.1] and is called the optimal coupling. The corresponding optimal value of (A.3) is referred to as the Wasserstein-2 distance

$$W_2(P, Q) = \left( \inf_{\pi \in \Pi(P, Q)} \int \|x - y\|^2 d\pi(x, y) \right)^{1/2}.$$

The above optimization problem is an (infinite-dimensional) convex program with linear constraints, and it admits a dual maximization problem, known as the Kantorovich dual problem given below.

$$W_2^2(P, Q) = \sup_{(\phi, \nu) \in \mathcal{K}} \int \phi dP + \int \nu dQ, \quad (\text{A.4})$$

where  $\mathcal{K}$  is the set of pairs  $(\phi, \nu) \in L^1(\Omega) \times L^1(\Omega)$  such that  $\phi(x) + \nu(y) \leq \|x - y\|^2$  for all  $x, y \in \Omega$ . Let  $\phi_0, \nu_0$  be the pair for which the supremum is achieved. The Kantorovich dual problem can be

reparameterized and it is equivalent to the following semi-dual problem

$$W_2^2(P, Q) = \sup_{\psi \in L^1(\Omega)} \int \psi dP + \int \psi^c dQ. \quad (\text{A.5})$$

The Brenier potential  $\psi$  is the solution of the semi-dual problem in (A.5) and is related to  $\phi_0$  as  $\psi = \|\cdot\|^2 - 2\phi_0$ . So far, we focus on the cost function  $c(x, y) = \|x - y\|^2$ , but we can define the optimal transport cost more generally as well. Given a non-negative cost function  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ , the *optimal transport cost* based on  $c$  is defined by

$$\text{OT}_c(P, Q) := \inf_{\pi \in \Pi(P, Q)} \int c(x, y) d\pi(x, y). \quad (\text{A.6})$$

Consider i.i.d. random variables  $X_1, X_2, \dots, X_n \sim P$  and  $Y_1, Y_2, \dots, Y_n \sim Q$  and let  $P_n$  and  $Q_n$  denote their corresponding empirical measures, i.e.  $P_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$  and  $Q_n = \frac{1}{n} \sum_{k=1}^n \delta_{Y_k}$ . The *empirical optimal transport cost* based on  $c$  is defined as

$$\text{OT}_c(P_n, Q_n) := \inf_{\sigma} \frac{1}{n} \sum_{k=1}^n c(X_k, Y_{\sigma(k)}), \quad \sigma \text{ is a permutation of } \{1, 2, \dots, n\}.$$

For the special case of  $c(x, y) = c_p(x, y) := \|x - y\|^p$  (note that the cost function  $c_p(x, y)$  is negative of the quality function  $\varphi_p(x, y)$ ) for  $p \geq 1$ . For  $p \in [1, \infty)$ , the Wasserstein  $p$ -distance between two probability distributions  $P$  and  $Q$  on Borel sets of  $\mathbb{R}^d$  with finite  $p$ -moments is defined as

$$W_p(P, Q) := \left( \inf_{\pi \in \mathcal{M}(P, Q)} \int \|x - y\|^p d\pi(x, y) \right)^{1/p} = \left( \inf_{\pi \in \mathcal{M}(P, Q)} \mathbb{E}_{(x, y) \sim \pi} [\|x - y\|^p] \right)^{1/p}. \quad (\text{A.7})$$

Note that  $W_p^p(P, Q) = \text{OT}_{c_p}(P, Q) = -U_{\infty}(P, Q)$ , where  $U_{\infty}(P, Q)$  is the fluid optimum for the quality function  $\varphi_p(x, y) = -\|x - y\|^p$ . Moreover, we have that

$$\text{OT}_{c_p}(P_n, Q_n) = W_p^p(P_n, Q_n) = \inf_{\sigma} \frac{1}{n} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p, \quad \sigma \text{ is a permutation over } \{1, 2, \dots, n\}.$$

Note that  $\mathbb{E} [W_p^p(P_n, Q_n)] = -U_n^H(P, Q)$  where  $U_n^H$  is the hindsight optimum matching value for the quality function  $\varphi_p(X, Y) = -\|X - Y\|^p$ .

### A.6.2 Existing Results on convergence of Empirical Optimal Transport value

In this section, we will present some results on the convergence of the empirical optimal transport value to its limit for different quality functions and different assumptions on the distributions  $P$  and  $Q$ . A function  $c : U \rightarrow \mathbb{R}$  on a convex domain  $U \subseteq \mathbb{R}^d$  is  $(\alpha, \Lambda)$ -Holder smooth for  $0 < \alpha \leq 1$  and  $\Lambda > 0$  if  $\|c\|_\infty < \Lambda$  and  $|c(x) - c(y)| \leq \Lambda \|x - y\|^\alpha$  and  $c$  is  $(\alpha, \Lambda)$ -Hölder smooth for  $1 < \alpha \leq 2$  if  $\|c\|_\infty \leq \Lambda$  and  $c$  is differentiable with  $(\alpha - 1, \Lambda)$ -Hölder smooth partial derivatives.

**Lemma 6 (Theorem 3.11 of [162])** Fix  $d \geq 1$  and a constant  $M < \infty$ . Let  $\mathcal{X}$  and  $\mathcal{Y}$  be Polish spaces and  $c : \mathcal{X} \times \mathcal{Y} \rightarrow [0, M]$  be continuous. If  $c$  is  $(\alpha, \Lambda)$ -Hölder smooth for some  $\alpha \in [1, 2]$ , then for any  $P \in \mathcal{P}(\mathcal{X})$  and  $Q \in \mathcal{P}(\mathcal{Y})$ , there exists a constant  $C \equiv C(P, Q, d, \alpha)$  such that

$$\mathbb{E} [|\text{OT}_c(P_n, Q_n) - \text{OT}(P, Q)|] \leq \begin{cases} Cn^{-1/2}, & \text{if } d < 2\alpha, \\ Cn^{-1/2} \log n, & \text{if } d = 2\alpha, \\ Cn^{-\alpha/d}, & \text{if } d > 2\alpha. \end{cases}$$

**Corollary 9** Fix  $d \geq 1$ . Let  $\mathcal{X}$  and  $\mathcal{Y}$  be bounded subsets of  $\mathbb{R}^d$  with a non-empty interior. Consider  $c_p(x, y) = \|x - y\|^p$  and assume that the demand  $P$  and supply  $Q$  distributions are supported on  $\mathcal{X}$  and  $\mathcal{Y}$  respectively, then there exists a constant  $C \equiv C(P, Q, d, p) < \infty$  such that

$$\mathbb{E} [W_p^p(P_n, Q_n) - W_p^p(P, Q)] \leq \begin{cases} Cn^{-1/2}, & \text{if } d < 2(p \wedge 2), \\ Cn^{-1/2} \log n, & \text{if } d = 2(p \wedge 2), \\ Cn^{-(p \wedge 2)/d}, & \text{if } d > 2(p \wedge 2). \end{cases}$$

*Proof of Corollary 9.* Since both  $\mathcal{X}$  and  $\mathcal{Y}$  are closed, bounded and convex, we have that  $\mathcal{X}$  and  $\mathcal{Y}$  are Polish spaces. Since  $\mathcal{X}$  and  $\mathcal{Y}$  are compact and  $c_p$  is continuous, we have that  $\|c_p\|_\infty \leq M$  for

some constant  $M < \infty$ . From [53, Corollary 3], we know that  $c_p$  is  $(p \wedge 2, \Lambda)$ -Hölder smooth, i.e.  $\alpha = p \wedge 2$ . Hence all the conditions in Lemma 6 are verified. From the convexity of  $f(x) = |x|$ , we have that

$$\mathbb{E} [W_p^p(P_n, Q_n) - W_p^p(P, Q)] \leq |\mathbb{E} [W_p^p(P_n, Q_n) - W_p^p(P, Q)]| \leq \mathbb{E} [|W_p^p(P_n, Q_n) - W_p^p(P, Q)|].$$

Finally, using the fact that  $\text{OT}_{c_p}(P_n, Q_n) = W_p^p(P_n, Q_n)$  and  $\text{OT}_{c_p}(P, Q) = W_p^p(P, Q)$ , the results follows from Lemma 6. ■

**Lemma 7 (Proposition 21 of [53])** *Fix  $d \geq 1$ .  $\mathcal{X}$  and  $\mathcal{Y}$  are convex subsets of  $\mathbb{R}^d$  with non-empty interior. The cost function  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$  and takes the form  $c(x, y) = h(x - y)$  where  $h : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$  is convex, even, and lower semi-continuous. Further assume that  $h$  is differentiable over  $\mathcal{Z} = \mathcal{X} - \mathcal{Y}$ . Furthermore, there exists  $\lambda > 0, \alpha \in (0, 2]$  and  $z_0 = x_0 - y_0 \in \mathcal{Z}$  such that  $x_0 \in \text{int}(\mathcal{X}), y_0 \in \text{int}(\mathcal{Y})$  and for all  $z \in \mathcal{Z}$ ,*

$$h(z) - h(z_0) \geq \begin{cases} \lambda \|z - z_0\|^\alpha, & \alpha \leq 1, \\ \langle \nabla h(z_0), z - z_0 \rangle + \lambda \|z - z_0\|^\alpha, & \alpha > 1. \end{cases}$$

Then there exists a constant  $c > 0$  such that

$$\sup_{P \in \mathcal{P}(\mathcal{X}), Q \in \mathcal{P}(\mathcal{Y})} \mathbb{E} [\text{OT}_c(P_n, Q_n) - \text{OT}_c(P, Q)] \geq cn^{-\alpha/d}.$$

**Corollary 10** *Fix  $d \geq 1$ . Let  $\mathcal{X}$  and  $\mathcal{Y}$  be closed, bounded and convex subsets of  $\mathbb{R}^d$  with a non-empty interior. Consider  $c_p(x, y) = \|x - y\|^p$ . There exists demand  $P$  and supply  $Q$  distributions supported on  $\mathcal{X}$  and  $\mathcal{Y}$  respectively and a constant  $c \equiv c(P, Q, d, p) > 0$  such that*

$$\mathbb{E} [W_p^p(P_n, Q_n) - W_p^p(P, Q)] \geq cn^{-(p \wedge 2)/d}.$$

*Proof of Corollary 10.* From [53], it is straightforward to see that  $h(x) = \|x\|^p$  satisfies the condi-

tions in Lemma 7 with  $\alpha = p \wedge 2$ . The conditions on  $\mathcal{X}$  and  $\mathcal{Y}$  are satisfied by assumption. The result follows from Lemma 7.  $\blacksquare$

**Lemma 8 ([58])** Fix  $d \geq 1$ . Let  $P = Q = \text{Uniform}([0, 1]^d)$ . Consider the cost function  $c_p(x, y) = \|x - y\|^p$  for some  $p \in [1, d]$ . Then we have that

$$\mathbb{E} [W_p^p(P_n, Q_n)] = \begin{cases} \Theta(n^{-p/2}), & d = 1, \\ \Theta((\log n)^{p/2} n^{-p/2}), & d = 2, \\ \Theta(n^{-p/d}), & d \geq 3. \end{cases}$$

**Lemma 9 (Proposition 13 of [60])** Let  $P$  and  $Q$  be absolutely continuous distributions with the support in  $[0, 1]^d$  and with densities being bounded above and below over  $[0, 1]^d$ . Furthermore assume that Assumption 1 is satisfied. Then for the cost function  $c(x, y) = \|x - y\|^2$ , we have that there exists a constant  $C \equiv C(P, Q, d) < \infty$  such that

$$\mathbb{E} [W_2^2(P_n, Q_n) - W_2^2(P, Q)] \leq \begin{cases} Cn^{-1}, & d = 1, \\ Cn^{-1} \log n, & d = 2, \\ Cn^{-2/d}, & d \geq 3. \end{cases}$$

### A.6.3 Equivalence of $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$ and $\varphi_2(X, Y) = -\|X - Y\|^2$

We formally establish that the regret corresponding to the dot-product quality function  $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$  scales exactly as the regret corresponding to the quality  $\varphi(X, Y) = -\|X - Y\|^2$ .

**Lemma 10** The dot-product quality function  $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$  and the quality function  $\varphi_2(X, Y) = -\|X - Y\|^2$  are equivalent in that the regret incurred by any policy  $\pi$  under  $\varphi(X, Y) = \langle X, Y \rangle$  is exactly half the regret under  $\varphi_2(X, Y) = -\|X - Y\|^2$ , incurred by the same  $\pi$ .

*Proof of Lemma 10.* Consider the quality function  $\varphi_2(X, Y) = -\|X - Y\|^2$ , then we have that  $U_\infty(P, Q, \varphi_2) = -\lim_{n \rightarrow \infty} \mathbb{E} [\inf_{\sigma \in \mathcal{S}_n} \sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2] / n$ . For any policy  $\pi \in \Pi$ , the expected

average quality under the policy  $\pi$  with quality function  $\varphi_2(X, Y) = -\|X - Y\|^2$  is given as  $U_n(\pi; P, Q, \varphi_2) = -\mathbb{E} \left[ \sum_{t=1}^n \|X_t - Y_{\pi(t)}\|^2 \right] / n$ .

Consider the hindsight optimum value for the dot product utility  $\varphi(X, Y) = \langle X, Y \rangle$  for  $n \geq 1$ ,

$$\begin{aligned} U_n^H(P, Q, \varphi_{\text{dot}}) &\stackrel{(a)}{=} \frac{1}{n} \mathbb{E} \left[ \sup_{\sigma \in S_n} \sum_{t=1}^n \langle X_t, Y_{\sigma(t)} \rangle \right] \stackrel{(b)}{=} \frac{1}{2n} \mathbb{E} \left[ \sup_{\sigma \in S_n} \sum_{t=1}^n \langle X_t, X_t \rangle + \langle Y_{\sigma(t)}, Y_{\sigma(t)} \rangle - \|X_t - Y_{\sigma(t)}\|^2 \right] \\ &\stackrel{(c)}{=} \frac{1}{2} \mathbb{E} [\langle X, X \rangle] + \frac{1}{2} \mathbb{E} [\langle Y, Y \rangle] - \frac{1}{2n} \mathbb{E} \left[ \inf_{\sigma \in S_n} \sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2 \right], \end{aligned}$$

where (a) follows from the definition of  $U_n^H(P, Q, \varphi_{\text{dot}})$  for  $\varphi_{\text{dot}}(x, y) = \langle x, y \rangle$ , (b) follows from the fact that  $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle = \frac{1}{2} \langle X, X \rangle + \frac{1}{2} \langle Y, Y \rangle - \frac{1}{2} \|X - Y\|^2$ , (c) from the fact that  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$  are i.i.d. Therefore we have that  $U_\infty(P, Q, \varphi_{\text{dot}}) = \mathbb{E} [\langle X, X \rangle] / 2 + \mathbb{E} [\langle Y, Y \rangle] / 2 - \lim_{n \rightarrow \infty} \mathbb{E} \left[ \inf_{\sigma \in S_n} \sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2 \right] / 2n$ . Hence we have that

$$U_\infty(P, Q, \varphi_{\text{dot}}) = \mathbb{E} [\langle X, X \rangle] / 2 + \mathbb{E} [\langle Y, Y \rangle] / 2 + U_\infty(P, Q, \varphi_2) / 2.$$

For any policy  $\pi$  and the quality function  $\varphi_{\text{dot}}(x, y) = \langle x, y \rangle$ , the expected average quality under the policy  $\pi$  is given as

$$\begin{aligned} U_n(\pi; P, Q, \varphi_{\text{dot}}) &= \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \langle X_t, Y_{\pi(t)} \rangle \right] = \frac{1}{2n} \mathbb{E} \left[ \sum_{t=1}^n \langle X_t, X_t \rangle + \langle Y_{\pi(t)}, Y_{\pi(t)} \rangle - \|X_t - Y_{\pi(t)}\|^2 \right] \\ &= \frac{1}{2} \mathbb{E} [\langle X, X \rangle] + \frac{1}{2} \mathbb{E} [\langle Y, Y \rangle] - \frac{1}{2n} \mathbb{E} \left[ \sum_{t=1}^n \|X_t - Y_{\pi(t)}\|^2 \right] \\ &= \frac{1}{2} \mathbb{E} [\langle X, X \rangle] + \frac{1}{2} \mathbb{E} [\langle Y, Y \rangle] + \frac{1}{2} U_n(\pi; P, Q, \varphi_2). \end{aligned}$$

Hence we have that for any policy  $\pi$ ,

$$\begin{aligned} \text{REG}_n(\pi; P, Q, \varphi_{\text{dot}}) &= U_\infty(P, Q, \varphi_{\text{dot}}) - U_n(\pi; P, Q, \varphi_{\text{dot}}) \\ &= U_\infty(P, Q, \varphi_2) / 2 - U_n(\pi; P, Q, \varphi_2) / 2 \\ &= \text{REG}_n(\pi; P, Q, \varphi_2) / 2. \end{aligned}$$

Hence we have that the regret for the quality functions  $\varphi_2(x, y)$  and  $\varphi_{\text{dot}}(x, y)$  are equal up to a constant factor and this completes the proof.  $\blacksquare$

## A.7 Proof of Theorem 2

*Proof of Theorem 2.* We begin by proving the upper bound on the performance of the SOAR algorithm. We have that  $U_n(\text{SOAR}) = -\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \|X_t - Y_{\pi_t^{\text{SOAR}}}\|^p\right]$  and  $U_\infty = -W_p^p(P, Q)$ . Using the definition of regret (per match) and Theorem 1, we have that

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &= U_\infty - U_n(\text{SOAR}) \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \|X_t - \pi_t^{\text{SOAR}}(X_t)\|^p\right] - W_p^p(P, Q), \\ &\stackrel{(a)}{=} \frac{1}{n}\sum_{k=1}^n \left(\frac{1}{k}\mathbb{E}\left[\min_{\sigma} \sum_{j=1}^k \|X_j - Y_{\sigma(j)}\|^p\right] - W_p^p(P, Q)\right) \\ &\stackrel{(b)}{=} \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left[W_p^p(P_k, Q_k) - W_p^p(P, Q)\right], \end{aligned}$$

where (a) follows from Theorem 1 and (b) follows from the definition of  $W_p^p(P_k, Q_k)$ . Next we will consider three cases: (i)  $d < 2(p \wedge 2)$ , (ii)  $d = 2(p \wedge 2)$  and (iii)  $d > 2(p \wedge 2)$ .

(i)  $d < 2(p \wedge 2)$ . Using Corollary 9 for  $d < 2(p \wedge 2)$ , there is a constant  $C \equiv C(P, Q, d, p) < \infty$  such that  $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{1}{2}}$  for any  $k \geq 1$ . Hence we have that,

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n}\sum_{k=1}^n Ck^{-\frac{1}{2}} \leq \frac{C}{n}\int_0^n x^{-\frac{1}{2}}dx = C'n^{-\frac{1}{2}}.$$

(ii)  $d = 2(p \wedge 2)$ . Using Corollary 9 for  $d = 2(p \wedge 2)$ , there is a constant  $C \equiv C(P, Q, d, p) < \infty$  such that  $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{1}{2}} \log k$  for any  $k \geq 2$ . Hence we have that,

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n}\left(C + \sum_{k=2}^n Ck^{-\frac{1}{2}} \log k\right) \leq \frac{C' \log n}{n}\int_0^n x^{-\frac{1}{2}}dx = 2C'n^{-\frac{1}{2}} \log n.$$

(iii)  $d > 2(p \wedge 2)$ . Using Corollary 9 for  $d = 2(p \wedge 2)$ , there is a constant  $C \equiv C(P, Q, d, p) < \infty$  such that  $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{p \wedge 2}{d}}$  for any  $k \geq 1$ . Hence we have that

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n} \sum_{k=1}^n Ck^{-\frac{p \wedge 2}{d}} \leq \frac{C}{n} \int_0^n x^{-\frac{p \wedge 2}{d}} dx = C'n^{-\frac{p \wedge 2}{d}}.$$

This completes the proof of the upper bound in Theorem 2.

Next we will prove the lower bound on the regret of any online optimal policy. For any feasible, non-anticipative online policy  $\pi$ , recall that  $U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$  for any pair of distributions  $P$  and  $Q$  and hence we have that

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) = \inf_{\pi \in \Pi} U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi) \geq U_\infty(P, Q, \varphi) - U_n^H(P, Q, \varphi).$$

Recall that for  $\varphi(X, Y) = -\|X - Y\|^p$ , we have that  $U_n^H(P, Q, \varphi) = -\frac{1}{n} \mathbb{E} [\min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p] = -\mathbb{E} [W_p^p(P_n, Q_n)]$  and  $U_\infty(P, Q, \varphi) = -W_p^p(P, Q)$  and hence we have that

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq \mathbb{E} [W_p^p(P_n, Q_n) - W_p^p(P, Q)]. \quad (\text{A.8})$$

As before we will consider different cases for  $d$  depending on the value of  $p$ .

(i)  $d > 2(p \wedge 2)$ . From Corollary 10, there exists a positive constant  $c := c(P, Q, d, p) > 0$  such that  $\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq cn^{-\frac{p \wedge 2}{d}}$  for  $d > 2(p \wedge 2)$ .

(ii)  $d \leq 2(p \wedge 2)$ . We will consider the case where  $P = Q$  and distribution  $P$  is supported on the vertices of the hypercube  $[0, 1]^d$  namely on the points  $\mathcal{V} = \{\mathbf{v} : \mathbf{v}_i \in \{0, 1\} \forall i \in \{1, 2, \dots, d\}\}$  and moreover, the distribution  $P$  is uniform over the points in  $\mathcal{V}$ . Since  $P = Q$ , we have that  $U_\infty = -\inf_{\mu \in \mathcal{M}(P, Q)} \mathbb{E}[\|X - \mu(Y)\|^p] = 0$ . We have  $n$  i.i.d. samples of the demand  $X_1, X_2, \dots, X_n$  and supply  $Y_1, Y_2, \dots, Y_n$ . Let  $N_{\mathbf{v}}^X = \sum_{k=1}^n \mathbb{1}\{X_k = \mathbf{v}\}$  denote the number of demand random variables (i.e.  $X_1, X_2, \dots, X_n$ ) that are equal to the point  $\mathbf{v}$ . Similarly define the quantity  $N_{\mathbf{v}}^Y$  for the supply random variables  $Y_1, Y_2, \dots, Y_n$ . For a fixed  $\mathbf{v} \in \mathcal{V}$ ,

consider the following events  $\mathcal{E}_X = \{N_{\mathbf{v}}^X \leq n/2^d - \sqrt{n/2^d}\}$  and  $\mathcal{E}_Y = \{N_{\mathbf{v}}^Y \geq n/2^d + \sqrt{n/2^d}\}$ . Since  $P$  and  $Q$  are uniformly distributed over the points in  $\mathcal{V}$ , for any  $\mathbf{v} \in \mathcal{V}$ , we have that  $N_{\mathbf{v}}^X \sim \text{Bin}(n, 1/2^d)$  and  $N_{\mathbf{v}}^Y \sim \text{Bin}(n, 1/2^d)$  and hence  $\mathbb{E}[N_{\mathbf{v}}^X] = \mathbb{E}[N_{\mathbf{v}}^Y] = n/2^d$  and  $\text{var}(N_{\mathbf{v}}^X) = \text{var}(N_{\mathbf{v}}^Y) = n(2^d - 1)/2^{2d}$ . Using CLT, it is easy to observe that there exists a positive constant  $\alpha > 0$  such that  $\mathbb{P}(\mathcal{E}_X) \geq \alpha$  and  $\mathbb{P}(\mathcal{E}_Y) \geq \alpha$  and since the events  $\mathcal{E}_X$  and  $\mathcal{E}_Y$  are independent, we have that  $\mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \geq \alpha^2 > 0$ . Now we have that

$$\begin{aligned}
\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) &\stackrel{(a)}{\geq} \frac{1}{n} \mathbb{E} \left[ \min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \right] \\
&\stackrel{(b)}{=} \frac{1}{n} \mathbb{E} \left[ \min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| \mathcal{E}_X \cap \mathcal{E}_Y \right] \mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \\
&\quad + \frac{1}{n} \mathbb{E} \left[ \min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| (\mathcal{E}_X \cap \mathcal{E}_Y)^c \right] \mathbb{P}((\mathcal{E}_X \cap \mathcal{E}_Y)^c) \\
&\stackrel{(c)}{\geq} \frac{1}{n} \mathbb{E} \left[ \min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| \mathcal{E}_X \cap \mathcal{E}_Y \right] \mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \\
&\stackrel{(d)}{\geq} 2\alpha^2 \sqrt{n/2^d} / n = cn^{-\frac{1}{2}},
\end{aligned}$$

where (a) follows from that  $U_{\infty} = 0$  (since  $P = Q$ ) and (A.8), (b) follows from law of total expectations, (c) follows from the fact that  $\mathbb{E}[\min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| (\mathcal{E}_X \cap \mathcal{E}_Y)^c] \geq 0$ , (d) follows from fact that under the event  $\mathcal{E}_X \cap \mathcal{E}_Y$ , the number of supply units equal to  $\mathbf{v}$  (recall that we fixed  $\mathbf{v}$  while defining  $\mathcal{E}_X$  and  $\mathcal{E}_Y$ ) are at least  $2\sqrt{n/2^d}$  times more than the demand units equal to  $\mathbf{v}$  and hence these excess supply units must be matched to some  $\mathbf{v}' \neq \mathbf{v}$  and since  $\|\mathbf{v}' - \mathbf{v}\|^p \geq 1$  for any  $\mathbf{v}' \neq \mathbf{v}$ , we have that  $\mathbb{E}[\min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| \mathcal{E}_X \cap \mathcal{E}_Y] \geq 2\sqrt{n/2^d}$  and also the fact that  $\mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \geq \alpha^2 > 0$ .

Together this completes the proof of Theorem 2. ■

## A.8 Proof of Proposition 2

*Proof of Proposition 2.* We will begin by proving the upper bound on the performance of the

SOAR algorithm. We assume that  $P = Q = \text{Uniform}([0, 1]^d)$  and hence we have that  $U_\infty = 0$  for  $\varphi_p(X, Y) = -\|X - Y\|^p$  for any  $p \geq 1$  and we have that  $U_n(\text{SOAR}) = -\frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \|X_t - Y_{\pi_t^{\text{SOAR}}}\|^p \right]$ . Using the definition of regret (per match) and Theorem 1, we have that

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &\stackrel{(a)}{=} U_\infty - U_n(\text{SOAR}), \\ &\stackrel{(b)}{=} \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \|X_t - Y_{\pi_t^{\text{SOAR}}}\|^p \right], \\ &\stackrel{(c)}{=} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[ \frac{1}{k} \min_{\sigma} \sum_{j=1}^k \|X_j - Y_{\sigma(j)}\|^p \right], \\ &\stackrel{(d)}{=} \frac{1}{n} \sum_{k=1}^n \mathbb{E} [W_p^p(P_k, Q_k)]. \end{aligned}$$

where (a) follows from definition of regret, (b) follows from  $U_n(\text{SOAR})$ , (c) follows from Theorem 1, (d) follows from the definition of  $W_p^p(P_k, Q_k)$ . Next we will consider three cases : (i)  $d = 1$ , (ii)  $d = 2$  and (iii)  $d \geq 3$ .

(i)  $d = 1$ . Using Lemma 8 for  $d = 1$ , there is a constant  $C \equiv C(P, Q, p) < \infty$  such that  $\mathbb{E} [W_p^p(P_k, Q_k)] \leq Ck^{-p/2}$  for  $k \geq 1$  and  $p \geq 1$ . Hence we have that,

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &\leq \frac{1}{n} \sum_{k=1}^n Ck^{-p/2} \\ &\leq \frac{C}{n} \left( 1 + \int_1^n x^{-p/2} dx \right) \\ &\leq C' \left( n^{-p/2} \mathbb{1}\{p < 2\} + n^{-1} \log n \mathbb{1}\{p = 2\} + n^{-1} \mathbb{1}\{p > 2\} \right). \end{aligned}$$

(ii)  $d = 2$ . Using Lemma 8 for  $d = 2$ , there is a constant  $C \equiv C(P, Q, p) < \infty$  such that

$\mathbb{E} [W_p^p(P_k, Q_k)] \leq C(\log k)^{p/2} k^{-p/2}$  for  $k \geq 2$  and  $p \geq 1$ . Hence we have that,

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &\leq \frac{1}{n} \left( C + \sum_{k=2}^n C(\log k)^{p/2} k^{-p/2} \right), \\ &\stackrel{(a)}{\leq} \frac{C'}{n} \left( (\log n)^{p/2} \int_0^n x^{-p/2} dx \right) \mathbb{1}\{p < 2\} + \left( \log n \int_1^n x^{-1} dx \right) \mathbb{1}\{p = 2\} \\ &\quad + \frac{C'}{n} \left( \int_1^n x^{-1-\epsilon(p)} dx \right) \mathbb{1}\{p > 2\}, \\ &\leq C' \left( (\log n)^{p/2} n^{-p/2} \mathbb{1}\{p < 2\} + n^{-1} \log^2 n \mathbb{1}\{p = 2\} + n^{-1} \mathbb{1}\{p > 2\} \right), \end{aligned}$$

where (a) for  $p \leq 2$  follows from the fact that  $(\log k)^{p/2} \leq (\log n)^{p/2}$  for all  $k \leq n$  and for  $p > 2$ , we can write  $p = 2 + 2\epsilon$  for some  $\epsilon > 0$  and there exists  $n_0 \in \mathbb{N}$  such that  $(\log n)^{p/2} \leq n^\epsilon$  for all  $n \geq n_0$  and hence we have that  $(\log x/x)^{p/2} \leq x^{-1-\epsilon(p)}$  for sufficiently large  $x$ .

(iii)  $d \geq 3$ . Using Lemma 8 for  $d = 3$ , there is a constant  $C \equiv C(P, Q, p) < \infty$  such that

$\mathbb{E} [W_p^p(P_k, Q_k)] \leq Ck^{-p/d}$  for  $k \geq 1$  and  $p \geq 1$ . Hence we have that,

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &\leq \frac{1}{n} \sum_{k=1}^n Ck^{-p/d} m \\ &\leq \frac{C}{n} \left( 1 + \int_1^n x^{-p/d} dx \right) \\ &\leq C' \left( n^{-p/d} \mathbb{1}\{p < d\} + n^{-1} \log n \mathbb{1}\{p = d\} + n^{-1} \mathbb{1}\{p > d\} \right). \end{aligned}$$

This completes the upper bound proof. Next we will prove the lower bound on the regret of any online optimal policy. For any feasible, non-anticipative online policy  $\pi$ , recall that  $U_n^H(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$  for any pair of distributions  $P$  and  $Q$  and hence we have that

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) = \inf_{\pi \in \Pi} U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi) \geq U_\infty(P, Q, \varphi) - U_n^H(P, Q, \varphi).$$

Recall that for  $\varphi(X, Y) = -\|X - Y\|^p$ , we have that  $U_n^H(P, Q, \varphi) = -\frac{1}{n} \mathbb{E} \left[ \min_{\sigma} \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \right] =$

$-\mathbb{E} [W_p^p(P_n, Q_n)]$  and  $U_\infty(P, Q, \varphi) = -W_p^p(P, Q) = 0$  since  $P = Q$  and hence we have that

$$\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq \mathbb{E} [W_p^p(P_n, Q_n)]. \quad (\text{A.9})$$

We will consider the following three cases: (i)  $d = 1$ , (ii)  $d = 2$  and (iii)  $d \geq 3$ .

(i)  $d = 1$ . For the case of  $p \leq 2$ , from Lemma 8 and (A.9) it follows that  $\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq cn^{-p/2}$  for some constant  $c \equiv c(P, Q, p) > 0$ .

For the case of  $p > 2$ , observe that  $n^{-1} \mathbb{E} [\sum_{t=1}^n |X_t - Y_{\pi_t}|^p] \geq n^{-1} \mathbb{E} [|X_n - Y_{\pi_n}|^p] \geq c/n$  for some constant  $c > 0$ , where the last inequality follows from the fact that  $\mathbb{E} [|X_n - Y_{\pi_n}|^p] \geq c$ . This is because of the following reason. At time  $n$ , we have only one supply unit remaining. Let  $B$  be a ball of radius  $r = \Gamma(d/2 + 1)^{1/d} / \sqrt{\pi} \cdot 2^{-1/d}$  around the location of the last supply unit. We have that  $\text{Vol}(B) \leq 1/2$ . Since  $\text{Vol}([0, 1]^d) = 1$ , with probability at least  $1/2$ , a demand unit arrives in  $[0, 1]^d \cap B^c$  and the distance of demand unit is at least  $r$ . Note that this is irrespective of the location of the supply unit and hence the expected matching cost at  $t = n$  is at least  $(r/2)^p$  using Jensen's inequality.

(ii)  $d = 2$ . For the case of  $p \leq 2$ , from Lemma 8 and (A.9) it follows that  $\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq c(\log n)^{p/2} n^{-p/2}$  for some constant  $c \equiv c(P, Q, p) > 0$ .

For the case of  $p > 2$ , observe that  $n^{-1} \mathbb{E} [\sum_{t=1}^n \|X_t - Y_{\pi_t}\|^p] \geq n^{-1} \mathbb{E} [\|X_n - Y_{\pi_n}\|^p] \geq c/n$  for some constant  $c > 0$ . The reason for this lower bound follows the exact same reason as in the case of  $d = 1$ .

(iii)  $d \geq 3$ . For the case of  $p < d$ , from Lemma 8 and (A.9) it follows that  $\inf_{\pi \in \Pi} \text{REG}_n(\pi; P, Q, \varphi) \geq cn^{-p/d}$  for some constant  $c \equiv c(P, Q, p) > 0$ .

For the case of  $p > d$ , observe that  $n^{-1} \mathbb{E} [\sum_{t=1}^n \|X_t - Y_{\pi_t}\|^p] \geq n^{-1} \mathbb{E} [\|X_n - Y_{\pi_n}\|^p] \geq c/n$  for some constant  $c > 0$ . The reason for this lower bound follows the exact same reason as in the case of  $d = 1$ .

Finally for the case of  $p = d$ , we have that

$$\sum_{t=1}^n \mathbb{E} [\|X_t - Y_{\pi_t}\|^p] \stackrel{(a)}{\geq} \sum_{t=1}^n (\mathbb{E} [\|X_t - Y_{\pi_t}\|])^p \stackrel{(b)}{\geq} c \sum_{t=1}^n (n-t+1)^{-1} \geq c \int_1^n x^{-1} dx \geq c' \log n,$$

where (a) follows from Jensen's inequality, (b) follows from the following argument: at decision epoch  $t$ , we have  $n-t+1$  supply units in the  $[0, 1]^d$ . For any arbitrary location of supply units, we have that the expected distance between the incoming demand unit and the closest supply unit is at least  $c(n-t+1)^{-\frac{1}{d}}$  and hence the expected matching cost is of the order  $c(n-t+1)^{-1}$ . Let  $B_i$  be a ball of radius  $r = 2c(n-t+1)^{-\frac{1}{d}}$  around the  $i$ -th supply unit for  $1 \leq i \leq n-t+1$ . The volume of the ball  $B_i$  is proportional to  $r^d = 2^d c^d (n-t+1)^{-1}$ . For  $c$  chosen small enough, we have that  $\text{Vol}(B_i) \leq \frac{1}{2(n-t+1)}$ . Define  $B = \cup_{i=1}^{n-t+1} B_i$ . Now we have that  $\text{Vol}(B) \leq \sum_{i=1}^{n-t+1} \text{Vol}(B_i) \leq \frac{1}{2}$ . Since  $\text{Vol}([0, 1]^d) = 1$ , with probability at least  $1/2$ , a demand unit arrives in  $[0, 1]^d \cap B^c$  and the distance of the demand unit is at least  $c(n-t+1)^{-\frac{1}{d}}$ . This implies that the expected matching cost is at least  $\Omega((n-t+1)^{-1})$ .

This completes the lower bound proof. ■

## A.9 Vanishing Regret for polynomial kernel quality function

In this section, we discuss how the performance guarantees for the dot-product quality function  $\varphi_{\text{dot}}(X, Y) = \langle X, Y \rangle$  can be leveraged to establish vanishing regret guarantees for the broad class of quality functions which we refer to as the polynomial kernel quality functions  $\varphi_{\text{ker}}(X, Y) = \sum_{q=0}^m a_q \langle X, Y \rangle^q$ .

**Corollary 11** *Suppose  $P$  and  $Q$  are supported on bounded sets with dimension  $d$ . Fix  $m \in \mathbb{N}$  and consider the quality function  $\varphi_{\text{ker}}(X, Y) = \sum_{q=0}^m a_q \langle X, Y \rangle^q$  where  $a_q \geq 0$  for all  $q \leq m$ . Define  $d' \triangleq \sum_{q=0}^m \binom{d+q-1}{q} \mathbb{1}\{a_q > 0\}$ . There exists a universal constant  $C := C(P, Q, d, \{a_q\}_{q=0}^m) < \infty$  such that*

$$\text{REG}_n(\text{SOAR}) \leq C \left( n^{-\frac{1}{2}} \mathbb{1}\{d' \leq 3\} + n^{-\frac{1}{2}} \log n \mathbb{1}\{d' = 4\} + n^{-\frac{2}{d'}} \mathbb{1}\{d' \geq 5\} \right)$$

*Proof of Corollary 11.* We first consider the quality function  $\varphi_{\text{ker}}^q(X, Y) = \langle X, Y \rangle^q$  for some  $q \in \mathbb{N}$ . From [62], it follows that there exists a continuous mapping  $\phi_q : \mathbb{R}^d \rightarrow \mathbb{R}^{d_q}$  where  $d_q \triangleq \binom{d+q-1}{q}$  such that

$$\varphi_{\text{ker}}^q(X, Y) = \langle X, Y \rangle^q = \langle \phi_q(X), \phi_q(Y) \rangle \quad (\text{A.10})$$

In the case that  $a_q > 0, q = 1, \dots, m$ , we define  $\phi_q$  in the following form:

$$\phi(z) \triangleq \begin{bmatrix} \sqrt{a_0} \\ \sqrt{a_1}\phi_1(z) \\ \sqrt{a_2}\phi_2(z) \\ \vdots \\ \sqrt{a_m}\phi_m(z) \end{bmatrix}_{d' \times 1},$$

In general, when there exists  $q$  s.t.  $a_q = 0$ , we simply remove the corresponding terms  $\sqrt{a_q}\phi_q(z)$  from the above RHS. Hence the dimension of  $\phi$  is  $d' = \sum_{q=0}^m \binom{d+q-1}{q} \mathbb{1}\{a_q > 0\}$ . Let  $\mathcal{X}$  and  $\mathcal{Y}$  denote the support of the distributions  $P$  and  $Q$  respectively. Define  $\mathcal{X}' = \{\phi(x) : x \in \mathcal{X}\} \subseteq \mathbb{R}^{d'}$  and  $\mathcal{Y}' = \{\phi(y) : y \in \mathcal{Y}\} \subseteq \mathbb{R}^{d'}$ . Let  $P'$  and  $Q'$  denote the resulting distributions on the set  $\mathcal{X}'$  and  $\mathcal{Y}'$  respectively. Using (A.10), we have that

$$\varphi_{\text{ker}}(X, Y) = \sum_{q=0}^m a_q \langle X, Y \rangle^q = \langle \phi(X), \phi(Y) \rangle = \langle X', Y' \rangle$$

Note that  $P'$  and  $Q'$  satisfy the assumptions of Corollary 3 since  $P$  and  $Q$  are supported on bounded sets and  $\phi$  is a continuous mapping. Therefore, the result in Corollary 11 follows by invoking Corollary 3. ■

### A.10 Proof of Theorem 3

*Proof of Theorem 3.* We begin by proving the upper bound on the regret for the SOAR al-

gorithm. From Lemma 10, we know that regret of SOAR for the dot-product quality function  $\varphi_{\text{dot}} = \langle X, Y \rangle$  is given as

$$\begin{aligned} \text{REG}_n(\text{SOAR}) &= \frac{1}{2} \left[ \frac{1}{n} \mathbb{E} \left[ \sum_{t=1}^n \|X_t - Y_{\pi_t^{\text{SOAR}}}\|^2 \right] - W_2^2(P, Q) \right] \\ &= \frac{1}{2n} \sum_{k=1}^n \mathbb{E} [W_2^2(P_k, Q_k) - W_2^2(P, Q)], \end{aligned}$$

where  $P_k$  and  $Q_k$  denote the empirical measure corresponding to  $k$  i.i.d samples from the distributions  $P$  and  $Q$  respectively. Next we consider the following three cases: (i)  $d = 1$ , (ii)  $d = 2$  and (iii)  $d \geq 3$ .

(i)  $d = 1$ . Using Lemma 9 for  $d = 1$ , there is a constant  $C \equiv C(P, Q) < \infty$  such that  $\mathbb{E}[W_2^2(P_k, Q_k) - W_2^2(P, Q)] \leq Ck^{-1}$  for all  $k \geq 1$ . Hence we have that

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n} \sum_{k=1}^n Ck^{-1} \leq \frac{C}{n} \left( 1 + \int_1^n x^{-1} dx \right) \leq C'n^{-1} \log n.$$

(ii)  $d = 2$ . Using Lemma 9 for  $d = 2$ , there is a constant  $C \equiv C(P, Q) < \infty$  such that  $\mathbb{E}[W_2^2(P_k, Q_k) - W_2^2(P, Q)] \leq Ck^{-1}(\log k)^2$  for any  $k \geq 2$ . Hence we have that,

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n} \left( C + \sum_{k=2}^n Ck^{-1}(\log k)^2 \right) \leq \frac{C'(\log n)^2}{n} \int_1^n x^{-1} dx = 2C'n^{-1}(\log n)^3.$$

(iii)  $d \geq 3$ . Using Lemma 9 for  $d \geq 3$ , there is a constant  $C \equiv C(P, Q) < \infty$  such that  $\mathbb{E}[W_2^2(P_k, Q_k) - W_2^2(P, Q)] \leq Ck^{-2/d}$  for any  $k \geq 1$ . Hence we have that

$$\text{REG}_n(\text{SOAR}) \leq \frac{1}{n} \sum_{k=1}^n Ck^{-\frac{2}{d}} \leq \frac{C}{n} \int_0^n x^{-\frac{2}{d}} dx = C'n^{-\frac{2}{d}}.$$

This completes the proof of the upper bound in Theorem 2. The lower bound follows from the lower bound in Proposition 2 for  $p = 2$ . ■

## Appendix B: Dynamic Resource Allocation: Algorithmic Design Principles and Spectrum of Achievable Performances

### B.1 Proof of Theorem 4

First we will consider the case of  $\beta = 0$ . For the uniform distribution over  $[0, 1]$ , we have that  $\beta = 0$  and from Proposition 4 of [70], Theorem 4 follows for  $\beta = 0$ . Therefore our focus will be on the case of  $\beta > 0$ . Fix  $\beta > 0$  and fix a number  $g \geq 0$ . In the context of Example 3, we have that  $g = \frac{1}{2}$ . Consider a distribution supported on the set  $\mathcal{S} \triangleq [0, \ell] \cup [u, 1]$  where  $\ell \triangleq \frac{1}{2} - \frac{g}{2}$  and  $u \triangleq \frac{1}{2} + \frac{g}{2}$ . For  $g = 1/2$ , we have that  $\ell = \frac{1}{4}$  and  $u = \frac{3}{4}$ . For a fixed  $\beta > 0$  and  $g, \ell, u$  as defined above, consider the following candidate ability distribution  $F_{\beta, \ell, u}$ ,

$$F_{\beta, \ell, u}(x) = \begin{cases} -\frac{(\ell-x)^{1+\beta}}{2\ell^{1+\beta}} + \frac{1}{2}, & 0 \leq x \leq \ell \\ \frac{1}{2}, & \ell \leq x \leq u \\ \frac{(x-u)^{1+\beta}}{2(1-u)^{1+\beta}} + \frac{1}{2}, & u \leq x \leq 1 \end{cases} \quad (\text{B.1})$$

For  $g > 0$ , we can easily verify that  $F_{\beta, \ell, u}$  is a  $(\beta, \varepsilon_0 = \frac{1}{2})$ -clustered distribution and for  $g = 0$ ,  $F_{\beta, \ell, u}$  is a  $(\beta, \varepsilon_0 = 1)$ -clustered distribution. Next, we will fix the time horizon  $T > 0$  and set the budget  $B \triangleq \lfloor \frac{1}{2}T \rfloor$ . Define  $\Delta_\beta \triangleq T^{-\frac{1}{2(1+\beta)}}(1-u) = T^{-\frac{1}{2(1+\beta)}}\ell$ . Define  $c_0 \triangleq \left(\frac{129}{128}\right)^{\frac{1}{1+\beta}} > 1$ . Define the following quantities:

$$\alpha_0 \triangleq c_0 - 1 > 0, \quad \tilde{\Delta}_\beta \triangleq \alpha_0 \Delta_\beta, \quad \ell_1 \triangleq \ell - \Delta_\beta, \quad \ell_2 \triangleq \ell - c_0 \Delta_\beta, \quad u_1 \triangleq u + \Delta_\beta, \quad u_2 \triangleq u + c_0 \Delta_\beta. \quad (\text{B.2})$$

Now we will partition the set  $\mathcal{S} \triangleq [0, \ell] \cup [u, 1]$  into the following sets (refer to Figure B.1):

$$\mathcal{I}_L = [0, \ell_2), \quad \mathcal{I}_{M_1} = [\ell_2, \ell_1), \quad \mathcal{I}_{M_2} = [\ell_1, \ell], \quad \mathcal{I}_{M_3} = [u, u_1), \quad \mathcal{I}_{M_4} = [u_1, u_2), \quad \mathcal{I}_H = [u_2, 1]$$

Further define the sets  $\mathcal{I}_{M_c} \triangleq \mathcal{I}_{M_2} \cup \mathcal{I}_{M_3}$  and  $\mathcal{I}_{M_p} \triangleq \mathcal{I}_{M_1} \cup \mathcal{I}_{M_4}$ .

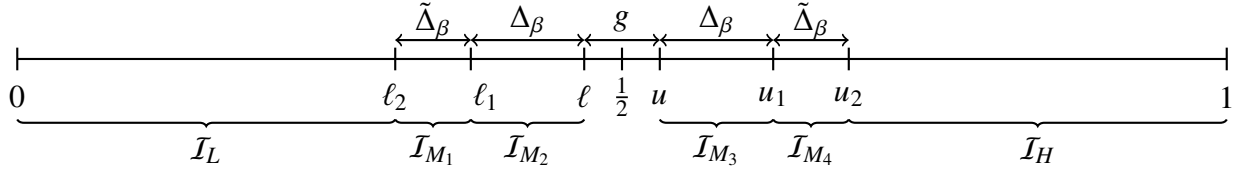


Figure B.1: Partition of the set  $\mathcal{S} = [0, \ell] \cup [u, 1]$  into disjoint set  $\mathcal{I}_L = [0, \ell_2)$ ,  $\mathcal{I}_{M_1} = [\ell_2, \ell_1)$ ,  $\mathcal{I}_{M_2} = [\ell_1, \ell]$ ,  $\mathcal{I}_{M_3} = [u, u_1)$ ,  $\mathcal{I}_{M_4} = [u_1, u_2)$ ,  $\mathcal{I}_H = [u_2, 1]$ , where  $\ell_1 \triangleq \ell - \Delta_\beta$ ,  $\ell_2 \triangleq \ell - c_0\Delta_\beta$ ,  $u_1 \triangleq u + \Delta_\beta$ ,  $u_2 \triangleq u + c_0\Delta_\beta$  and  $\tilde{\Delta}_\beta \triangleq (c_0 - 1)\Delta_\beta$ .

Let  $\theta_{\geq 1}$  denote the sequence of candidate abilities and define  $N(\mathcal{A}, t_1, t_2)$  denote the number of candidate abilities in the set  $\mathcal{A}$  that arrive in the time interval  $[t_1, t_2]$ . Formally, the random variable  $N(\mathcal{A}, t_1, t_2)$  is defined as

$$N(\mathcal{A}, t_1, t_2) \triangleq \sum_{k=t_1}^{t_2} \mathbb{1}\{\theta_k \in \mathcal{A}\}, \quad \forall \mathcal{A} \subseteq \mathcal{S}, t_1, t_2 \in \{1, 2, \dots, T\} \quad (\text{B.3})$$

Let  $\mu_{t_1}^{t_2}(\mathcal{A}) \triangleq \mathbb{E}[N(\mathcal{I}_H, t_1, t_2)]$  denote the mean of the random variable  $N(\mathcal{I}_H, t_1, t_2)$ .

Next we define the following set of events:

$$\mathcal{H}_1 \triangleq \left\{ \frac{T}{4} - \frac{\sqrt{T}}{2} \leq N(\mathcal{I}_H, 1, B) \leq \frac{T}{4} \right\} \quad (\text{B.4})$$

$$\mathcal{H}_2 \triangleq \left\{ \frac{T}{4} - 4\sqrt{T} \leq N(\mathcal{I}_H, B+1, T) \leq \frac{T}{4} - 3\sqrt{T} \right\}, \quad (\text{B.5})$$

$$\tilde{\mathcal{H}}_2 \triangleq \left\{ \frac{T}{4} + \frac{\sqrt{T}}{2} \leq N(\mathcal{I}_H, B+1, T) \leq \frac{T}{4} + \frac{3\sqrt{T}}{2} \right\}, \quad (\text{B.6})$$

$$\mathcal{C}_1 \triangleq \left\{ \frac{\sqrt{T}}{4} \leq N(\mathcal{I}_{M_c}, 1, B) \leq \sqrt{T} \right\}, \quad (\text{B.7})$$

$$\mathcal{C}_2 \triangleq \left\{ \frac{\sqrt{T}}{4} \leq N(\mathcal{I}_{M_c}, B+1, T) \leq \sqrt{T} \right\}, \quad (\text{B.8})$$

$$\mathcal{P}_1 \triangleq \left\{ \frac{\sqrt{T}}{256} \leq N(\mathcal{I}_{M_p}, 1, B) \leq \frac{\sqrt{T}}{64} \right\}, \quad (\text{B.9})$$

$$\mathcal{P}_2 \triangleq \left\{ \frac{\sqrt{T}}{256} \leq N(\mathcal{I}_{M_p}, B+1, T) \leq \frac{\sqrt{T}}{64} \right\}. \quad (\text{B.10})$$

Further we define the events  $\mathcal{H} \triangleq \mathcal{H}_1 \cap \mathcal{H}_2$ ,  $\tilde{\mathcal{H}} \triangleq \mathcal{H}_1 \cap \tilde{\mathcal{H}}_2$ ,  $\mathcal{C} \triangleq \mathcal{C}_1 \cap \mathcal{C}_2$  and  $\mathcal{P} \triangleq \mathcal{P}_1 \cap \mathcal{P}_2$ .

**Discussion of the Hindsight Optimal.** Conditional on the event  $\tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$ , we have that the total number of arrivals in the set  $\mathcal{I}_H$  is more than the budget  $B$  i.e.  $N(\mathcal{I}_H, 1, T) \geq \frac{1}{2}T \geq B$  and hence the hindsight optimal must reject *all* the arrivals in the set  $\mathcal{I}_L \cup \mathcal{I}_{M_p} \cup \mathcal{I}_{M_c}$  and *possibly some* arrivals in the set  $\mathcal{I}_H$ . However, conditional on the event  $\mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , we have that total number of arrivals in the set  $\mathcal{I}_H \cup \mathcal{I}_{M_c} \cup \mathcal{I}_{M_p}$  is less than the budget  $B$  i.e.  $N(\mathcal{I}_H, 1, T) + N(\mathcal{I}_{M_c}, 1, T) + N(\mathcal{I}_{M_p}, 1, T) \leq \frac{1}{2}T - \frac{31}{32}\sqrt{T} < \lfloor \frac{1}{2}T \rfloor = B$  for sufficiently large  $T$  and hence the hindsight optimal must accept *all* the arrivals in the set  $\mathcal{I}_H \cup \mathcal{I}_{M_p} \cup \mathcal{I}_{M_c}$  and *possibly some* arrivals in the set  $\mathcal{I}_L$ .

Let  $N^{\text{DP}}(\mathcal{A}, t_1, t_2)$  denote the number of accepted candidates by the DP (optimal dynamic programming policy) with ability in the set  $\mathcal{A}$  and they arrive in the time interval  $[t_1, t_2]$  which we formally define as:

$$N^{\text{DP}}(\mathcal{A}, t_1, t_2) \triangleq \sum_{k=t_1}^{t_2} \mathbb{1}\{\theta_k \in \mathcal{A}, \pi_k^{\text{DP}} = \text{accept}\}, \quad \forall \mathcal{A} \subseteq \mathcal{S}, t_1, t_2 \in \{1, 2, \dots, T\} \quad (\text{B.11})$$

Define the event  $\mathcal{E}$  which says that under the optimal online policy, the number of accepted candidates up till time  $B(= \lfloor \frac{1}{2}T \rfloor)$  is at least one eighth of the number of arrivals in set  $\mathcal{I}_{M_c}$  up till time  $B$ , i.e.,

$$\mathcal{E} \triangleq \left\{ N^{\text{DP}}(\mathcal{I}_{M_c}, 1, B) \geq \frac{N(\mathcal{I}_{M_c}, 1, B)}{8} \right\} \quad (\text{B.12})$$

**Proof Strategy.** Our proof will proceed by considering the following events: (a)  $\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$  and (b)  $\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ . In case (a), from the discussion about the hindsight optimal policy, the hindsight optimal policy will reject all the arrivals in the set  $\mathcal{I}_{M_c}$  but DP will accept at least  $\frac{1}{32}\sqrt{T}$  arrivals in set  $\mathcal{I}_{M_c}$  in the time interval  $[1, B]$ . This will result in the DP *incorrectly* rejecting at least  $\frac{1}{32}\sqrt{T}$  arrivals in interval  $\mathcal{I}_H$  and the cost of each of these mistakes is at least  $\alpha_0\Delta_\beta$ . In case (b), from the discussion about the hindsight optimal policy, the hindsight optimal policy will accept all the arrivals in the interval  $\mathcal{I}_{M_c}$  but the DP accepts at most  $\frac{1}{8}\sqrt{T}$  arrivals in the interval  $\mathcal{I}_{M_c}$ . This implies that at least  $\frac{1}{8}\sqrt{T}$  arrivals in the interval  $\mathcal{I}_{M_c}$  are *incorrectly* rejected. This will result in the DP *incorrectly* accepting at least  $\frac{1}{8}\sqrt{T}$  arrivals in the interval  $\mathcal{I}_L$  and the cost of each of these mistakes is again at least  $\alpha_0\Delta_\beta$ . Informally speaking, we can lower bound the expected regret as

$$\text{Regret}(B, T; \text{DP}) \geq c \left( \mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}) \right) [(\# \text{ of mistakes}) \times (\text{cost/mistake})]$$

Assuming we can show that  $\mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}) \geq \gamma > 0$  for some  $\gamma \in (0, 1)$ , we have that # of mistakes are  $\Omega(\sqrt{T})$  and cost of each mistake is  $\Omega\left(T^{-\frac{1}{2(1+\beta)}}\right)$ . Combining all this will provide the lower bound guarantee as desired for  $\beta > 0$ .

Consider the random variable  $\Lambda(B, T; \text{DP})$

$$\Lambda(B, T; \text{DP}) = \sum_{t=1}^T \theta_t a_t^{\text{hs}} - \sum_{t=1}^T \theta_t a_t^{\text{DP}} \quad (\text{B.13})$$

Next we will formalize our proof strategy using the following two lemmas.

**Lemma 11** Consider the event  $\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$ , then we have that

$$\mathbb{E} [\Lambda(B, T; DP) | \mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}] \geq \frac{\alpha_0 \ell}{32} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}},$$

where  $\alpha_0 = \left(\frac{129}{128}\right)^{\frac{1}{1+\beta}} - 1$  defined in (B.2) and  $\ell = \frac{1}{2} - \frac{\delta}{2}$ .

**Lemma 12** Consider the event  $\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , then we have that

$$\mathbb{E} [\Lambda(B, T; DP) | \mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}] \geq \frac{\alpha_0}{8} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}},$$

where  $\alpha_0 = \left(\frac{129}{128}\right)^{\frac{1}{1+\beta}} - 1$  defined in (B.2).

We defer the proofs of Lemmas 11 and 12 to Sections B.1.1 and B.1.2 respectively. Finally, we have that

$$\begin{aligned} \text{Regret}(B, T; DP) &\stackrel{(a)}{=} \mathbb{E} [\Lambda(B, T; DP)], \\ &\stackrel{(b)}{\geq} \mathbb{E} [\Lambda(B, T; DP) | \mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}] \mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}) \\ &\quad + \mathbb{E} [\Lambda(B, T; DP) | \mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}] \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}), \\ &\stackrel{(c)}{\geq} \frac{\alpha_0}{32} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}} \left( \mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}) \right), \end{aligned} \quad (\text{B.14})$$

where (a) follows from the definition of (expected) regret, (b) follows from total law of expectations, (c) follows from Lemmas 11 and 12.

Observe that  $\tilde{\mathcal{H}} = \mathcal{H}_1 \cap \tilde{\mathcal{H}}_2$ ,  $\mathcal{H} = \mathcal{H}_1 \cap \mathcal{H}_2$ ,  $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$  and  $\mathcal{P} = \mathcal{P}_1 \cap \mathcal{P}_2$  and moreover  $\mathcal{C}_1 \perp \mathcal{C}_2$ ,  $\mathcal{P}_1 \perp \mathcal{P}_2$  and  $\mathcal{H}_1 \perp \mathcal{H}_2$ ,  $\tilde{\mathcal{H}}_2$  since the events  $\mathcal{H}_1, \mathcal{C}_1, \mathcal{P}_1$  only depend on the arrivals in the time interval  $[1, B]$  i.e.  $\{\theta_k\}_{k=1}^B$  whereas the events  $\mathcal{H}_2, \tilde{\mathcal{H}}_2, \mathcal{C}_2$  and  $\mathcal{P}_2$  only depend on the arrivals in the time interval  $[B+1, T]$  i.e.  $\{\theta_k\}_{k=B+1}^T$  and the arrivals by assumption are *i.i.d.* Additionally, the events  $\mathcal{E}, \mathcal{E}^c$  also only depend on the arrivals in the interval  $[1, B]$  and hence are independent

of  $\mathcal{H}_2, \tilde{\mathcal{H}}_2, \mathcal{C}_2$  and  $\mathcal{P}_2$ . Therefore, we have that

$$\begin{aligned}
& \mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}) \\
& \stackrel{(a)}{=} \mathbb{P}(\mathcal{E} \cap \mathcal{H}_1 \cap \tilde{\mathcal{H}}_2 \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{P}_1 \cap \mathcal{P}_2) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H}_1 \cap \mathcal{H}_2 \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{P}_1 \cap \mathcal{P}_2) \\
& \stackrel{(b)}{=} \mathbb{P}(\mathcal{E} \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) \mathbb{P}(\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) \mathbb{P}(\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2) \\
& \stackrel{(c)}{\geq} \min \left\{ \mathbb{P}(\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2) \right\} \cdot (\mathbb{P}(\mathcal{E} \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1)), \\
& \stackrel{(d)}{=} \min \left\{ \mathbb{P}(\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2) \right\} \mathbb{P}(\mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1), \tag{B.15}
\end{aligned}$$

where (a) follows from the definition of  $\mathcal{H}, \tilde{\mathcal{H}}, \mathcal{C}$  and  $\mathcal{P}$ , (b) follows from the fact that  $\mathcal{E} \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1 \perp \tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2$  and  $\mathcal{E}^c \cap \mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1 \perp \mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2$  using the arguments presented previously, (c) follows trivially, (d) follows from the law of total probability.

Now it suffices to show to that there exists a constant  $\alpha > 0$  independent of  $T$  such that for all  $T$  sufficiently large, we have that  $\mathbb{P}(\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) \geq \alpha$ . Using a CLT argument, one can easily see that  $\mathbb{P}(\mathcal{H}_1), \mathbb{P}(\tilde{\mathcal{H}}_2), \mathbb{P}(\mathcal{H}_2) \geq \alpha' > 0$  and  $\mathbb{P}(\mathcal{C}_1), \mathbb{P}(\mathcal{C}_2), \mathbb{P}(\mathcal{P}_1), \mathbb{P}(\mathcal{P}_2) \geq \alpha'$  for  $\alpha'' \neq \alpha'$ . However the events  $\mathcal{H}_1, \mathcal{C}_1, \mathcal{P}_1$  (similarly  $\mathcal{H}_2, \mathcal{C}_2, \mathcal{P}_2$  and  $\tilde{\mathcal{H}}_2, \mathcal{C}_2, \mathcal{P}_2$ ) are correlated and hence proving  $\mathbb{P}(\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) \geq \alpha$  requires a conditioning argument which we will illustrate now. We will argue this the event  $\mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1$  and the exact same argument works for the events  $\tilde{\mathcal{H}}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2$  and  $\mathcal{H}_2 \cap \mathcal{C}_2 \cap \mathcal{P}_2$ . We have that

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_1 \cap \mathcal{C}_1 \cap \mathcal{P}_1) & \stackrel{(a)}{=} \mathbb{P}(\mathcal{C}_1 \cap \mathcal{P}_1) \mathbb{P}(\mathcal{H}_1 | \mathcal{C}_1 \cap \mathcal{P}_1), \\
& \stackrel{(b)}{=} \mathbb{P}(\mathcal{H}_1) - \mathbb{P}(\mathcal{H}_1 | (\mathcal{C}_1 \cap \mathcal{P}_1)^c) \mathbb{P}((\mathcal{C}_1 \cap \mathcal{P}_1)^c), \\
& \stackrel{(c)}{\geq} \mathbb{P}(\mathcal{H}_1) - (\mathbb{P}(\mathcal{C}_1^c) + \mathbb{P}(\mathcal{P}_1^c)),
\end{aligned}$$

where (a) follows from the definition of conditional probability, (b) follows from the law of total probability i.e.  $\mathbb{P}(\mathcal{H}_1) = \mathbb{P}(\mathcal{H}_1 | \mathcal{C}_1 \cap \mathcal{P}_1) \mathbb{P}(\mathcal{C}_1 \cap \mathcal{P}_1) + \mathbb{P}(\mathcal{H}_1 | (\mathcal{C}_1 \cap \mathcal{P}_1)^c) \mathbb{P}((\mathcal{C}_1 \cap \mathcal{P}_1)^c)$  and (c) follows from the fact that  $\mathbb{P}(\mathcal{H}_1 | (\mathcal{C}_1 \cap \mathcal{P}_1)^c) \mathbb{P}((\mathcal{C}_1 \cap \mathcal{P}_1)^c) \leq \mathbb{P}((\mathcal{C}_1 \cap \mathcal{P}_1)^c) \leq \mathbb{P}(\mathcal{C}_1^c) +$

$\mathbb{P}(\mathcal{P}_1^c)$  where the first inequality follows from the fact that  $\mathbb{P}(\mathcal{H}_1 | (C_1 \cap \mathcal{P}_1)^c) \leq 1$  and the second inequality follows from the union bound. Using the exact same arguments we have that

$$\begin{aligned}\mathbb{P}(\mathcal{H}_2 \cap C_2 \cap \mathcal{P}_2) &\geq \mathbb{P}(\mathcal{H}_2) - (\mathbb{P}(C_2^c) + \mathbb{P}(\mathcal{P}_2^c)), \\ \mathbb{P}(\tilde{\mathcal{H}}_2 \cap C_2 \cap \mathcal{P}_2) &\geq \mathbb{P}(\tilde{\mathcal{H}}_2) - (\mathbb{P}(C_2^c) + \mathbb{P}(\mathcal{P}_2^c)).\end{aligned}$$

Next we present a few lemmas which would imply that  $\mathbb{P}(\tilde{\mathcal{H}}_2 \cap C_2 \cap \mathcal{P}_2), \mathbb{P}(\mathcal{H}_2 \cap C_2 \cap \mathcal{P}_2), \mathbb{P}(\tilde{\mathcal{H}}_2 \cap C_2 \cap \mathcal{P}_2) \geq 0.001$ .

**Lemma 13** *There exists  $T_0 < \infty$  such that for all  $T \geq T_0$ , we have that  $\mathbb{P}(\mathcal{H}_1), \mathbb{P}(\mathcal{H}_2), \mathbb{P}(\tilde{\mathcal{H}}_2) \geq 0.003$*

**Lemma 14** *There exists  $T_0 < \infty$  such that for all  $T \geq T_0$ , we have that  $\mathbb{P}(C_1^c), \mathbb{P}(C_2^c), \mathbb{P}(\mathcal{P}_1^c), \mathbb{P}(\mathcal{P}_2^c) \leq 0.001$*

We defer the proofs of Lemma 13 and 14 to Appendix B.1.3 and B.1.4 respectively. Using Lemmas 13 and 14 and (B.15), we have that  $\mathbb{P}(\mathcal{E} \cap \tilde{\mathcal{H}} \cap C \cap \mathcal{P}) + \mathbb{P}(\mathcal{E}^c \cap \mathcal{H} \cap C \cap \mathcal{P}) \geq 10^{-6}$ , combined with (B.14) concludes the proof. ■

### B.1.1 Proof of Lemma 11

Recall the definition of the random variable  $\Lambda(B, T; \text{DP}) = \sum_{k=1}^T \theta_k \pi_k^{\text{hs}} - \sum_{k=1}^T \theta_k$ . For a sequence of candidate ability arrivals  $\theta_{\geq 1}$ , we can define the following random set of indices  $\mathcal{J}^{\text{hs}}$  and  $\mathcal{J}^{\text{DP}}$  as

$$\mathcal{J}^{\text{hs}}(\mathcal{A}) \triangleq \{k : \theta_k \in \mathcal{A} \text{ and } \pi_k^{\text{hs}} = 1\}, \quad \mathcal{J}^{\text{DP}}(\mathcal{A}) \triangleq \{k : \theta_k \in \mathcal{A} \text{ and } \pi_k^{\text{DP}} = 1\}, \quad \forall \mathcal{A} \subseteq \mathcal{S} \tag{B.16}$$

Notice that we can equivalently write the sum of values chosen under the hindsight optimal and the DP policy as

$$\sum_{t=1}^T \theta_t \pi_t^{\text{hs}} = \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_L)} \theta_k + \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p})} \theta_k + \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c})} \theta_k + \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_H)} \theta_k \quad (\text{B.17})$$

$$\sum_{t=1}^T \theta_t \pi_t^{\text{DP}} = \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_L)} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_H)} \theta_k \quad (\text{B.18})$$

Now conditional on the event  $\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$ , we have that  $\mathcal{J}^{\text{hs}}(\mathcal{I}_L) = \mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p}) = \mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) = \emptyset$  and  $|\mathcal{J}^{\text{hs}}(\mathcal{I}_H)| = B$  and we have that  $|\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \geq \frac{1}{32} \sqrt{T}$ . This follows from the fact under the event  $\mathcal{E}$ , the DP accepts at least  $\frac{1}{8} N(\mathcal{I}_{M_c}, 1, B)$  and from the event  $\mathcal{C}_1$ , it follows that  $N(\mathcal{I}_{M_c}, 1, B) \geq \frac{1}{4} \sqrt{T}$ . Using this we have that

$$\sum_{t=1}^T \theta_t \pi_t^{\text{hs}} = \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_H)} \theta_k = \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_H)} \theta_k \quad (\text{B.19})$$

Since any online policy can select at most  $B$  candidates and the offline policy will select the top  $B$  candidates, we have that conditional on the event  $\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$ ,

$$|\mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| \geq |\mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_L)| \quad (\text{B.20})$$

Conditional on the event  $\mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}$ , we have that

$$\begin{aligned}
& \mathbb{E} [\Lambda(B, T; \text{DP}) | \mathcal{E} \cap \tilde{\mathcal{H}} \cap \mathcal{C} \cap \mathcal{P}] \\
& \stackrel{(a)}{=} \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)} \theta_k - \left( \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_L)} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})} \theta_k \right), \\
& \stackrel{(b)}{\geq} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| (u + c_0 \Delta_\beta) - \left( \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_L)} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})} \theta_k \right), \\
& \stackrel{(c)}{\geq} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| (u + c_0 \Delta_\beta) - \left( |\mathcal{J}^{\text{DP}}(\mathcal{I}_L)| \ell + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| (u + c_0 \Delta_\beta) + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| (u + \Delta_\beta) \right), \\
& \stackrel{(d)}{\geq} |\mathcal{J}^{\text{DP}}(\mathcal{I}_L)| [(u + c_0 \Delta_\beta) - \ell] + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| [(u + c_0 \Delta_\beta) - (u + c_0 \Delta_\beta)] \\
& \quad + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| [(u + c_0 \Delta_\beta) - (u + \Delta_\beta)] \\
& \stackrel{(e)}{=} |\mathcal{J}^{\text{DP}}(\mathcal{I}_L)| [(u + c_0 \Delta_\beta) - \ell] + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \alpha_0 \Delta_\beta, \\
& \stackrel{(f)}{\geq} \frac{\alpha_0 \ell}{32} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}},
\end{aligned}$$

where (a) follows from (B.18) and (B.19), (b) follows from the fact that  $\sum_{k \in \mathcal{S}} a_k \geq |\mathcal{S}| \min_{k \in \mathcal{S}} \{a_k\}$  and by construction, for all the arrivals in the set  $\mathcal{I}_H$ ,  $\theta_k \geq u + c_0 \Delta_\beta$ , (c) follows similar to (b), (d) follows from (B.20), (e) follows from the definition of  $\alpha_0$  in (B.2), (f) follows from the fact that  $|\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \geq \frac{1}{32} \sqrt{T}$  due to the event  $\mathcal{E}$  and from the fact that  $|\mathcal{J}^{\text{DP}}(\mathcal{I}_L)| \geq 0$ .  $\blacksquare$

### B.1.2 Proof of Lemma 12

Recall the definitions of set of indices  $\mathcal{J}^{\text{hs}}$  and  $\mathcal{J}^{\text{DP}}$  from (B.16) and decomposition of the sum of values chosen under hindsight optimal and the DP policy as given in (B.17) and (B.18). Recall from the discussion of the hindsight optimal that under the event  $\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , the hindsight optimal will accept all the candidates with abilities in the set  $\mathcal{I}_H, \mathcal{I}_{M_p}, \mathcal{I}_{M_c}$  and possibly

some candidates in the set  $\mathcal{I}_L$ . Conditional on the event  $\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , we have that,

$$\begin{aligned}
B &\stackrel{(a)}{=} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_L)| \\
&\stackrel{(b)}{=} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| \\
&\quad + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_L)|, \\
&\stackrel{(c)}{\geq} |\mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_L)|, \\
&\stackrel{(d)}{=} |\mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| + |\mathcal{J}^{\text{DP}}(\mathcal{I}_L) \setminus \mathcal{J}^{\text{hs}}(\mathcal{I}_L)| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_L)|,
\end{aligned}$$

where (a) follows from the fact the hindsight optimal will accept exactly  $B$  candidates, (b) follows from the fact that for countable set  $A, B$  such that  $B \subseteq A$ , we have that  $|A| = |A \setminus B| + |B|$ , (c) follows from the fact that any online policy will accept at most  $B$  candidates, (d) follows for the same reason as (b). This implies the following inequality,

$$|\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \geq |\mathcal{J}^{\text{DP}}(\mathcal{I}_L) \setminus \mathcal{J}^{\text{hs}}(\mathcal{I}_L)| \tag{B.21}$$

Conditional on the event  $\mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , we have that

$$\mathbb{E} [\Lambda(B, T; \text{DP}) | \mathcal{E}^c \cap \mathcal{H} \cap \mathcal{C} \cap \mathcal{P}]$$

$$\begin{aligned}
&\stackrel{(a)}{=} \left( \sum_{\mathcal{A} \in \{\mathcal{I}_H, \mathcal{I}_{M_p}, \mathcal{I}_{M_c}\}} \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{A}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{A})} \theta_k + \sum_{\mathcal{A} \in \{\mathcal{I}_H, \mathcal{I}_{M_p}, \mathcal{I}_{M_c}\}} \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{A})} \theta_k + \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_L)} \theta_k \right) \\
&\quad - \left( \sum_{\mathcal{A} \in \{\mathcal{I}_H, \mathcal{I}_{M_p}, \mathcal{I}_{M_c}\}} \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{A})} \theta_k + \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_L) \setminus \mathcal{J}^{\text{hs}}(\mathcal{I}_L)} \theta_k + \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{I}_L)} \theta_k \right), \\
&\stackrel{(b)}{=} \sum_{\mathcal{A} \in \{\mathcal{I}_H, \mathcal{I}_{M_p}, \mathcal{I}_{M_c}\}} \sum_{k \in \mathcal{J}^{\text{hs}}(\mathcal{A}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{A})} \theta_k - \sum_{k \in \mathcal{J}^{\text{DP}}(\mathcal{I}_L) \setminus \mathcal{J}^{\text{hs}}(\mathcal{I}_L)} \theta_k, \\
&\stackrel{(c)}{\geq} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| (u + c_0 \Delta_\beta) + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| (\ell - c_0 \Delta_\beta) \\
&\quad + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| (\ell - \Delta_\beta) - |\mathcal{J}^{\text{DP}}(\mathcal{I}_L) \setminus \mathcal{J}^{\text{hs}}(\mathcal{I}_L)| (\ell - c_0 \Delta_\beta) \\
&\stackrel{d}{\geq} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| [(u + c_0 \Delta_\beta) - (\ell - c_0 \Delta_\beta)] + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| [(\ell - \Delta_\beta) - (\ell - c_0 \Delta_\beta)] \\
&\quad + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_p}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_p})| [(\ell - c_0 \Delta_\beta) - (\ell - c_0 \Delta_\beta)] \\
&\stackrel{(e)}{=} |\mathcal{J}^{\text{hs}}(\mathcal{I}_H) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_H)| [(u + c_0 \Delta_\beta) - (\ell - c_0 \Delta_\beta)] + |\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \alpha_0 \Delta_\beta, \\
&\stackrel{(f)}{\geq} \frac{\alpha_0 \ell}{8} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}},
\end{aligned}$$

where (a) follows from (B.17) and (B.18), (b) follows trivially, (c) follows from the fact  $\theta_k \mathbb{1}\{\theta_k \in \mathcal{I}_H\} \geq u + c_0 \Delta_\beta$ ,  $\theta_k \mathbb{1}\{\theta_k \in \mathcal{I}_{M_p}\} \geq \ell - c_0 \Delta_\beta$ ,  $\theta_k \mathbb{1}\{\theta_k \in \mathcal{I}_{M_c}\} \geq \ell - \Delta_\beta$  and  $\theta_k \mathbb{1}\{\theta_k \in \mathcal{I}_L\} \leq \ell - c_0 \Delta_\beta$ , (d) follows from (B.21), (e) follows from the definition of  $\alpha_0 = c_0 - 1$ , (f) follows from the fact that  $|\mathcal{J}^{\text{hs}}(\mathcal{I}_{M_c}) \setminus \mathcal{J}^{\text{DP}}(\mathcal{I}_{M_c})| \geq \frac{1}{8} \sqrt{T}$  which is due to fact that under the event  $\mathcal{H} \cap \mathcal{C} \cap \mathcal{P}$ , the hindsight optimal will accept all the arrivals in the set  $\mathcal{I}_{M_c}$  however under the event  $\mathcal{E}^c$  will accept at most  $\frac{1}{8} \sqrt{T}$  arrivals in the first  $B$  time steps and this will result in incorrectly rejecting at least  $\frac{1}{4} \sqrt{T} - \frac{1}{8} \sqrt{T} = \frac{1}{8} \sqrt{T}$  arrivals in the set  $\mathcal{I}_{M_c}$ .  $\square$

### B.1.3 Proof of Lemma 13

Recall the definition of the events  $\mathcal{H}_1$ ,  $\mathcal{H}_2$  and  $\tilde{\mathcal{H}}_2$  as defined in (B.4), (B.5) and (B.6) respectively. We have that  $N(\mathcal{I}_H, 1, B) \sim \text{Bin}(B, p_{\mathcal{I}_H})$ , where  $p_{\mathcal{I}_H} = \frac{1}{2} - \frac{129}{256} T^{-\frac{1}{2}}$ . Therefore we have that

$\mu_1^B(\mathcal{I}_H) = \mathbb{E}[N(\mathcal{I}_H, 1, B)] \approx \frac{T}{4} - \frac{129}{512}\sqrt{T}$  since  $B \approx T/2$ . Therefore we can write the event  $\mathcal{H}_1$  as

$$\mathcal{H}_1 = \left\{ \mu_1^B(\mathcal{I}_H) - \frac{127}{512}\sqrt{T} \leq N(\mathcal{I}_H, 1, B) \leq \mu_1^B(\mathcal{I}_H) + \frac{129}{512}\sqrt{T} \right\}$$

Therefore we have that

$$\begin{aligned} \mathbb{P}(\mathcal{H}_1) &\stackrel{(a)}{=} \mathbb{P}\left(\mu_1^B(\mathcal{I}_H) - \frac{127}{512}\sqrt{T} \leq N(\mathcal{I}_H, 1, B) \leq \mu_1^B(\mathcal{I}_H) + \frac{129}{512}\sqrt{T}\right) \\ &\stackrel{(b)}{=} \mathbb{P}\left(-\frac{127}{512}\sqrt{T} \leq N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H) \leq \frac{129}{512}\sqrt{T}\right) \\ &\stackrel{(c)}{=} \mathbb{P}\left(-\frac{127}{256} \frac{1}{\sqrt{p_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{129}{256} \frac{1}{\sqrt{p_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}}\right) \\ &\stackrel{(d)}{\geq} \mathbb{P}\left(0 \leq \frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{258}{256}\right) \\ &\stackrel{(e)}{=} \mathbb{P}\left(\frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{258}{256}\right) - \mathbb{P}\left(\frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq 0\right) \\ &\stackrel{(f)}{\geq} \Phi\left(\frac{258}{256}\right) - \Phi(0) - \frac{c}{\sqrt{T}} \\ &\stackrel{(g)}{\geq} 0.34 - \frac{c}{\sqrt{T}} \end{aligned}$$

where (a) follows from definition of event  $\mathcal{H}_1$ , (b,c) follows trivially, (d) follows the fact that  $\left\{0 \leq \frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{258}{256}\right\} \subseteq \left\{-\frac{127}{256} \frac{1}{\sqrt{p_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{N(\mathcal{I}_H, 1, B) - \mu_1^B(\mathcal{I}_H)}{\sqrt{Bp_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}} \leq \frac{129}{256} \frac{1}{\sqrt{p_{\mathcal{I}_H}(1-p_{\mathcal{I}_H})}}\right\}$  since  $p_{\mathcal{I}_H}(1-p_{\mathcal{I}_H}) \leq 1/4$ , (e) follows trivially, (f) follows from Berry Esseen Theorem and (g) follows trivially. Now there exists a  $T_0 < \infty$  such that for all  $T \geq T_0$ , we have that  $\mathbb{P}(\mathcal{H}_1) \geq 0.003$ . An analogous proof follows for  $\mathcal{H}_2$  and  $\tilde{\mathcal{H}}_2$  as well, we omit it to avoid repetition.  $\square$

#### B.1.4 Proof of Lemma 14

Recall the definition of events  $C_1, C_2, \mathcal{P}_1$  and  $\mathcal{P}_2$  as defined in (B.7), (B.8), (B.9) and (B.10) respectively. We have that  $N(\mathcal{I}_{M_c}, 1, B) \sim \text{Bin}(B, p_{\mathcal{I}_{M_c}})$  where  $p_{\mathcal{I}_{M_c}} = T^{-\frac{1}{2}}$ . Therefore we have

that  $\mu_1^B(\mathcal{I}_{M_c}) = \mathbb{E} [N(\mathcal{I}_{M_c}, 1, B)] \approx \sqrt{T}/2$  since  $B \approx T/2$ . Now the event  $C_1$  can be written as

$$C_1 = \left\{ \mu_1^B(\mathcal{I}_{M_c})/2 \leq N(\mathcal{I}_{M_c}, 1, B) \leq 2\mu_1^B(\mathcal{I}_{M_c}) \right\}$$

Therefore we have that

$$\begin{aligned} \mathbb{P}(C_1^c) &\stackrel{(a)}{=} \mathbb{P}\left(\{N(\mathcal{I}_{M_c}, 1, B) \geq 2\mu_1^B(\mathcal{I}_{M_c})\} \cup \{N(\mathcal{I}_{M_c}, 1, B) \leq \mu_1^B(\mathcal{I}_{M_c})/2\}\right) \\ &\stackrel{(b)}{\leq} \mathbb{P}\left(\{N(\mathcal{I}_{M_c}, 1, B) \geq 2\mu_1^B(\mathcal{I}_{M_c})\}\right) + \mathbb{P}\left(\{N(\mathcal{I}_{M_c}, 1, B) \leq \mu_1^B(\mathcal{I}_{M_c})/2\}\right) \\ &\stackrel{(c)}{\leq} cT^{-\frac{1}{4}} \end{aligned}$$

where (a) follows from definition of  $C_1$ , (b) follows from union bound, (c) follows from Berry Esseen theorem as applied before. From this it follows that there exists  $T_0 < \infty$  such that  $\mathbb{P}(C_1^c) \leq 0.001$  for all  $T \geq T_0$ . An analogous proof follows for  $C_2, \mathcal{P}_1$  and  $\mathcal{P}_2$ , we omit it to avoid repetition.

□

## B.2 Details and Analysis of CWG Policy

In this section we will provide some more details about the CWG algorithm (Algorithm 2) and also provide the proof of Theorem 5. In Section B.2.1, we provide a discussion about the phase structure of Algorithm 2. In Section B.2.2, we define the concept of *hindsight-to-go* (HTG) which will aid our analysis. In Section B.2.3, we provide a proof outline for Theorem 5. In Section B.2.4, we will provide some helper lemmas to formalize our analysis with their proofs deferred to Section B.2.6. In Section B.2.5, we provide the formal proof of Theorem 5.

### B.2.1 Phase Structure of Algorithm 2

The phase structure of the CWG policy has been devised to simplify the analysis of the CWG policy. The key idea of the CWG policy is that if the CE threshold  $p_t^{\text{CE}}$  at time  $t$  is within a ball of radius  $\Delta_t \triangleq \sqrt{2 \log \tau / \tau}$  (where  $\tau = T - t + 1$  is the number of remaining time steps) of a gap

quantile  $q_i^*$ , then the CWG threshold is set to the gap quantile  $q_i^*$  itself. As  $t$  increases, so does the size of the radius and hence eventually there will be more than one gap quantiles in this ball. If there are more than one gap quantiles in  $\Delta_t$ , we need a tie-breaking rule to decide which gap quantile the CWG threshold is assigned to. This tie-breaking rule further complicates an already involved analysis and hence to simplify the technical analysis, we define the CWG algorithm by dividing it into two phases.

In the first phase, it suffices to ensure that there will always be at most one gap quantile in the  $\Delta_t$ -neighbourhood of  $p_t^{\text{CE}}$  for any value of  $p_t^{\text{CE}}$  and there is no need for a tie-breaking rule. One way to ensure this, is to find  $t^*$  such that for all  $t \leq t^*$ , we have that  $\sqrt{2 \log \tau / \tau} \leq \varepsilon_0 / 2$ . Note that irrespective of the value of  $p_t^{\text{CE}}$ , there is at most one gap quantile in the  $\Delta_t$  neighborhood of  $p_t^{\text{CE}}$ . Further, note that for  $t \leq T - 2$ ,  $\sqrt{2 \log \tau / \tau}$  is increasing in  $t$  and hence it suffices to verify that  $\sqrt{2 \log \tau^* / \tau^*} \leq \varepsilon_0 / 2$  for  $\tau^* = \lceil 64 \log(1/\varepsilon_0) / \varepsilon_0^2 \rceil$ . Given that we are guaranteed to have at most one gap quantile in  $\Delta_t$ -neighbourhood of  $p_t^{\text{CE}}$ , our analysis is greatly simplified.

The second phase is of length  $\lceil 64 \log(1/\varepsilon_0) / \varepsilon_0^2 \rceil$  and we use a static allocation rule in the second phase. The contribution to regret because of the static policy is at most  $C \sqrt{\log(1/\varepsilon_0) / \varepsilon_0}$  for some universal constant  $C < \infty$ .

### B.2.2 Hindsight To Go (HTG) and HTG Threshold

Let  $q_{\geq t}^\theta(n)$  denote the  $n$ -th largest value quantile in  $q_{\geq t}^\theta$  for an integer  $n \in \mathbb{N}$ . Define the following quantile values  $q_t^l \triangleq q_{\geq t}^\theta(B_t + 1)$  and  $q_t^u \triangleq q_{\geq t}^\theta(B_t)$  and denote their corresponding values by  $l_t = F^{-1}(q_t^l)$ ,  $u_t = F^{-1}(q_t^u)$ , where  $B_t$  is the remaining budget at time  $t$ . Note that since the principle of compensated coupling is to persuade the hindsight policy to take the same action as the online policy using sufficient compensations, the hindsight policy at time  $t$  may look different from the hindsight policy initially and being adapted to the budget which evolves according to the online policy. To distinguish between the two, at any time  $t$ , we will instead refer to the hindsight policy as the *Hindsight To Go* (HTG) policy, which due to coupling follows the same actions as the online policy up till time  $t - 1$  and then from time  $t$  onwards takes the optimal hindsight decision

with arrivals in  $\omega_{\geq t}$  given the remaining budget  $B_t$ . Given the CWG quantile threshold  $p_t^{\text{CWG}}$ , we define  $p_t^{\text{HTG}} \triangleq \arg \max_{x \in [q_t^l, q_t^u]} |p_t^{\text{CWG}} - x|$  when  $B_t > 0$ , otherwise  $p_t^{\text{HTG}} = 1$ . The reason to adopt this particular  $p_t^{\text{CWG}}$  dependent definition of  $p_t^{\text{HTG}}$  is that the compensation needed at time  $t$  will now be bounded above by the separation between the CWG threshold and the HTG threshold in value space.

### B.2.3 Proof Outline

We first provide a proof outline. Recall  $\tilde{T} = T - \lceil 64 \log(1/\varepsilon_0) \rceil / \varepsilon_0^2$  in Algorithm 2 (the CWG policy). The algorithm operates in two phases, the first phase includes time steps  $t$  such that  $1 \leq t \leq \tilde{T}$  while the second phase consists of the remaining time steps  $t$  such that  $\tilde{T} + 1 \leq t \leq T$ .

**Analysis of First Phase.** The analysis of the first phase makes use of the regret decomposition given in Lemma 1. To bound the expected compensation term  $\mathbb{E}_{B_t^\pi} [\partial \mathcal{R}_t(B_t^\pi, a_t^\pi)]$  in Lemma 1 for  $\pi = \text{CWG}$ , we will analyse two thresholds: the CWG quantile threshold denoted as  $p_t^{\text{CWG}}$  and Hindsight To Go (HTG) quantile threshold  $p_t^{\text{HTG}}$ . Note that given a tail sequence  $\theta_{\geq t}$  and the remaining budget  $B_t$ , the Hindsight To Go threshold is set such that on the sample path  $\theta_{\geq t}$ , the top  $B_t$  candidates are chosen. We bound the expected compensation at time  $t$  for  $t \in [1, \tilde{T}]$  and we do so by dividing the analysis into two events: (a)  $E_t = \{1 - B_t/\tau > 4\sqrt{\log \tau/\tau}\}$  and (b)  $E_t^c = \{1 - B_t/\tau \leq 4\sqrt{\log \tau/\tau}\}$  where  $\tau = T - t + 1$ . At any time either of the two events arises and we bound the expected compensation conditional on each of the two events. The analysis for both the events utilizes the same recipe. We show that with high probability the difference between CWG quantile threshold  $p_t^{\text{CWG}}$  and the HTG quantile threshold  $p_t^{\text{HTG}}$  is bounded above by  $C\sqrt{\log \tau/\tau}$  (Lemma 17). As a result of this, we establish that with high probability the two thresholds  $p_t^{\text{CWG}}$  and  $p_t^{\text{HTG}}$  belong to the same cluster (Lemma 18). Now compensation is need at time  $t$  only if there is a candidate ability arrival  $\theta_t$  such that its quantile  $F(\theta_t)$  lies between the two thresholds  $p_t^{\text{CWG}}$  and  $p_t^{\text{HTG}}$  and the amount of compensation is bounded by  $|F^{-1}(p_t^{\text{CWG}}) - F^{-1}(p_t^{\text{HTG}})|$  (Lemma 19). Using Lemmas 17, 18, 19 and definition of the  $(\beta, \varepsilon_0, \delta)$ -clustered distribution, we show that the

expected compensation at time  $t$  is bounded as follow.

**Lemma 15** *There is a universal constant  $C < \infty$  such that the following occurs. For any  $\beta \in [0, \infty)$ ,  $\varepsilon_0 \in (0, 1]$  and  $\delta \in (0, 1]$ , suppose the candidate-ability distribution  $F$  with associated gaps is  $(\beta, \varepsilon_0, \delta)$ -clustered. Then for  $t \in \{1, 2, \dots, \tilde{T}\}$ , for the CWG policy we have that the expected compensation at time  $t$  is bounded above as*

$$\sup_{B_t \geq 0} \partial \mathcal{R}_t \left( B_t, a_t^{\text{CWG}} \right) \leq C \left( (\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} + \delta \sqrt{\log \tau / \tau} \right),$$

where  $\tau = T - t + 1$ . Note that the above implies that

$$\mathbb{E}_{B_t} \left[ \partial \mathcal{R}_t \left( B_t, a_t^{\text{CWG}} \right) \right] \leq C \left( (\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} + \delta \sqrt{\log \tau / \tau} \right).$$

Using Lemma 1 and Lemma 15, the cummulative regret accrued up till time  $\tilde{T}$  is upper bounded by  $C \left( (\log T)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}} \mathbb{1}\{\beta > 0\} + \log^2 T \mathbb{1}\{\beta = 0\} + \delta \sqrt{T \log T} \right)$ .

**Analysis of Second Phase.** Recall that the CWG policy (Algorithm 2) in the last  $64 \log(1/\varepsilon_0)/\varepsilon_0^2$  time steps, makes use of the static allocation policy where we solve for the CE quantile threshold  $p_{\tilde{T}}^{\text{CE}}$  and thereafter use the time invariant quantile threshold  $p_{\tilde{T}}^{\text{CE}}$ . Using a well known fact in the network revenue management literature, we know that the regret accrued under a static allocation policy is upper bounded as  $C\sqrt{\text{horizon length}}$  for some universal constant  $C < \infty$ . Since the CWG policy (Algorithm 2) employs the static allocation policy for the last  $\lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$ , the regret accrued over the last  $T - \tilde{T}$  time steps is upper bounded as  $C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0$ . Adding up the regret over the two phases results in the regret scaling in Theorem 5.

#### B.2.4 Preliminaries and Helper Lemmas

We introduce some helper lemmas which we will use to prove the regret bound. We defer the proof of these lemmas to Appendix B.2.6. Let  $t$  denote the current time step and  $\tau = T - t + 1$  denote the remaining number of times steps. Assume that  $T \geq \lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$  and define

$\tilde{T} \triangleq T - \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$ . Define the following events for  $t \leq \tilde{T}$ :

$$\mathcal{A}_{1,t} = \{|p_t^{\text{CE}} - p_t^{\text{HTG}}| \leq \sqrt{2 \log \tau / \tau}\}, \quad (\text{B.22})$$

$$\mathcal{A}_{2,t} = \{|p_t^{\text{CWG}} - p_t^{\text{HTG}}| \leq 3\sqrt{\log \tau / \tau}\}, \quad (\text{B.23})$$

$$\mathcal{A}_{3,t} = \cup_{i=1}^{n+1} \{p_t^{\text{CWG}} \in \bar{Q}_i, p_t^{\text{HTG}} \in \bar{Q}_i\}, \quad (\text{B.24})$$

where  $\bar{Q}_i = [q_{i-1}^*, q_i^*]$  and  $n$  denotes the number of gaps. The interpretation of  $\mathcal{A}_{3,t}$  is that the CWG policy threshold and the HTG policy threshold belong (weakly) to the same mass cluster. The following lemmas show that these three events are very likely to occur for  $t \leq \tilde{T}$ :

**Lemma 16** *Consider the event  $\mathcal{A}_{1,t}$  defined in (B.22). We have that  $\mathbb{P}(\mathcal{A}_{1,t}^c) \leq 2/\tau^4$ .*

**Lemma 17** *Consider the event  $\mathcal{A}_{2,t}$  defined in (B.23). We have that  $\mathbb{P}(\mathcal{A}_{2,t}^c) \leq 2/\tau^4$ .*

**Lemma 18** *Consider the event  $\mathcal{A}_{3,t}$  defined in (B.24). We have that  $\mathbb{P}(\mathcal{A}_{3,t}^c) \leq 2n(n+1)/\tau^4$ , where  $n$  is the number of gaps.*

Let  $q_t^\theta = F(\theta_t)$  be the quantile of the candidate ability  $\theta_t$  at time  $t$ . If  $p_t^{\text{CWG}} < q_t^l$  then we have that  $p_t^{\text{HTG}} = q_t^u$  and compensation is needed only if  $q_t^\theta \in [p_t^{\text{CWG}}, p_t^{\text{HTG}}]$ . If  $p_t^{\text{CWG}} > q_t^u$  then we have that  $p_t^{\text{HTG}} = q_t^l$  and compensation is needed only if  $q_t^\theta \in [p_t^{\text{HTG}}, p_t^{\text{CWG}}]$ . If  $p_t^{\text{CWG}} \in (q_t^l, q_t^u)$ , then no compensation is required.

**Lemma 19** *Let  $q_t^\theta = F(\theta_t)$  denote the quantile corresponding to  $\theta_t$ . Compensation needs to be provided only if  $q_t^\theta \in (\min\{p_t^{\text{CWG}}, p_t^{\text{HTG}}\}, \max\{p_t^{\text{CWG}}, p_t^{\text{HTG}}\})$ ; let  $\partial R_t(B_t, a_t^{\text{CWG}})$  denote the compensation. Then we have that  $\partial R_t(B_t, a_t^{\text{CWG}}) \leq \max\{F^{-1}(p_t^{\text{CWG}}) - F^{-1}(p_t^{\text{HTG}}), F^{-1}(p_t^{\text{HTG}}) - F^{-1}((p_t^{\text{CWG}})^+)\}$ .*

### B.2.5 Formal Proof of Theorem 5

*Proof of Theorem 5.* Define  $\tilde{T} = T - \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$  and define  $\tau_0 = \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$ . Consider some time  $t \leq \tilde{T}$  and let  $\tau = T - t + 1$  denote the remaining time. Recall that  $p_t^{\text{CWG}} \in \mathcal{F}_t$  and

$p_t^{\text{HTG}}$  depends on the candidate abilities  $\theta_{\geq t}$  but only via the  $B_t$ -th largest quantile  $q_t^u$  and  $B_t + 1$ -th largest quantile  $q_t^l$ . To facilitate our analysis, we employ the so-called principle of deferred decisions, and only reveal  $q_t^u$  and  $q_t^l$  (in addition to the history up to time  $t$  i.e.  $\mathcal{F}_t$ ), which uniquely determines  $p_t^{\text{HTG}}$ . Define the event  $\mathcal{L}_t \triangleq \{p_t^{\text{CWG}} \leq q_t^l\}$  and  $\mathcal{H}_t \triangleq \{p_t^{\text{CWG}} \geq q_t^u\}$ . For the rest of the proof, we will condition on the event  $\mathcal{L}_t$  and prove an upper bound on the expected compensation  $\partial \mathcal{R}_t(B_t, a_t)$  (conditional on  $\mathcal{L}_t$ ). A similar bound can be analogously shown under the event  $\mathcal{H}_t$  and we omit the details to avoid repetition. Let  $q_t^\theta$  denote the quantile corresponding to the candidate ability  $\theta_t$ . Now compensation is needed only if  $q_t^\theta \in [p_t^{\text{CWG}}, q_t^u]$ . Let  $C_t$  denote the event that compensation is needed i.e., the action under the CWG threshold is different from the action under the HTG threshold. Given  $q_t^l$  and  $q_t^u$ , we know that the  $\tau$  periods to go include a random subset of  $B_t$  quantiles located above  $q_t^u$  (these quantiles are i.i.d uniform in  $[q_t^u, 1]$ ) and the remaining  $\tau - B_t$  quantiles are below  $q_t^l$  (these quantiles are i.i.d uniform in  $[0, q_t^l]$ ). If  $p_t^{\text{CWG}} \in (q_t^l, q_t^u)$ , no compensation is needed. Compensation is needed only if  $q_t^\theta \in [p_t^{\text{CWG}}, q_t^l]$  and this event occurs if (a) the realized quantile  $q_t^\theta = q_t^l$  or (b)  $q_t^\theta \in [p_t^{\text{CWG}}, q_t^l]$  is one of the  $\tau - B_t - 1$  lower quantiles. The probability of case (a) is  $1/\tau$  and probability of (b) is  $(\tau - B_t - 1)(q_t^l - p_t^{\text{CWG}})/q_t^l \tau$ . Combining the two we have that

$$\mathbb{P}\left(C_t | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t\right) = \frac{\mathbb{1}_{\{p_t^{\text{CWG}} \leq q_t^l\}}}{\tau} + \frac{(\tau - B_t - 1)(q_t^l - p_t^{\text{CWG}})_+}{q_t^l \tau} \quad (\text{B.25})$$

where  $\tau = T - t + 1$  and  $(x)_+ = \max\{x, 0\}$ . Using Lemma 19 and (B.25), we have the following bound the expected compensation

$$\mathbb{E}\left[\partial \mathcal{R}_t(B_t^{\text{CWG}}, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t\right] \leq \frac{|F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})_+)|}{\tau} + \frac{(q_t^l - p_t^{\text{CWG}})|F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})_+)|(\tau - B_t - 1)}{q_t^l \tau} \quad (\text{B.26})$$

Next we need to bound the ratio  $(\tau - B_t - 1)/(q_t^l \tau)$  and at any time  $t \leq T - \tau_0$ , exactly of

the following complementary events occurs: (a)  $\mathcal{E}_t = \{1 - B_t/\tau > 4\sqrt{\log \tau/\tau}\}$  and (b)  $\mathcal{E}_t^c = \{1 - B_t/\tau \leq 4\sqrt{\log \tau/\tau}\}$ . Recall the event  $\mathcal{A}_{3,t}$  defined in B.24, which states that

$$\mathcal{A}_{3,t} = \{q_t^l, q_t^u, p_t^{\text{CWG}} \text{ are quantiles belonging (weakly) to the same cluster}\}$$

Next we will establish an upper bound on (B.26) for each of the events  $\mathcal{E}_t$  and  $\mathcal{E}_t^c$ .

**Case (a):**  $\mathcal{E}_t = \{1 - B_t/\tau > 4\sqrt{\log \tau/\tau}\}$ . Define the following events:

$$\mathcal{A}_{4,t} \triangleq \{q_t^l \geq (1/2)(1 - B_t/\tau)\},$$

$$\mathcal{A}_{5,t} \triangleq \mathcal{A}_{3,t} \cap \mathcal{A}_{4,t}.$$

Under the event  $\mathcal{A}_{3,t}$ , from Definition 2 (a) it follows that  $|F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)| \leq |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{1+\beta}} + \delta$ . Now, on the event  $\mathcal{A}_{4,t}$ , we have that  $(\tau - B_t - 1)/(\tau q_t^l) \leq 2$ . We have that

$$\begin{aligned} \mathbb{P}(\mathcal{A}_{4,t}^c | B_t, \mathcal{E}_t) &= \mathbb{P}(q_t^l < (1/2)(1 - B_t/\tau)) \leq \mathbb{P}(\text{Binomial}(\tau, (1/2)(1 - B_t/\tau)^-) \geq \tau - B_t - 1) \\ &\leq \exp(-\Omega(\tau - B_t)) \leq C/(\tau - B_t)^4 \leq C/\tau^2. \end{aligned}$$

where the last inequality follows from the case assumption that  $\tau - B_t \geq 4\sqrt{\tau \log \tau}$  and the inequality is true for some appropriately defined constant  $C < \infty$ . It follows that

$$\mathbb{P}(\mathcal{A}_{4,t}^c | \mathcal{E}_t) \leq C/\tau^2. \tag{B.27}$$

Using (B.26), and the definitions of the events  $\mathcal{A}_{3,t}$  and  $\mathcal{A}_{4,t}$ , we have that

$$\begin{aligned} & \mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t, \mathcal{E}_t \right] \\ & \leq \mathbb{1}_{\mathcal{A}_{3,t}} \cdot \left[ |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{1+\beta}} + \delta \right] / \tau + 2 \mathbb{1}_{\mathcal{A}_{3,t}} \mathbb{1}_{\mathcal{A}_{4,t}} \left[ |q_t^u - p_t^{\text{CWG}}|^{1+\frac{1}{1+\beta}} + |q_t^u - p_t^{\text{CWG}}| \delta \right] + \mathbb{1}_{\mathcal{A}_{3,t}^c} + \mathbb{1}_{\mathcal{A}_{4,t}^c}, \end{aligned} \quad (\text{B.28})$$

$$\leq |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{1+\beta}} / \tau + \delta / \tau + 2 |q_t^u - p_t^{\text{CWG}}|^{1+\frac{1}{1+\beta}} + 2 |q_t^u - p_t^{\text{CWG}}| \delta + \mathbb{1}_{\mathcal{A}_{4,t}^c} + \mathbb{1}_{\mathcal{A}_{3,t}^c}, \quad (\text{B.29})$$

where the first inequality follows from  $q_t^l - p_t^{\text{CWG}} \leq q_t^u - p_t^{\text{CWG}}$ , and the second inequality follows from the fact that  $\mathbb{1}_{\mathcal{A}_{3,t}}, \mathbb{1}_{\mathcal{A}_{4,t}} \leq 1$ . Using the definition of the event  $\mathcal{A}_{2,t}$  in (B.23) and Lemma 17, we have that for all  $\alpha \in (0, 2]$ , we have

$$\mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^\alpha \right] \leq \mathbb{E} \left[ \mathbb{1}_{\mathcal{A}_{2,t}} |q_t^u - p_t^{\text{CWG}}|^\alpha + \mathbb{1}_{\mathcal{A}_{2,t}^c} \right] \leq 3^\alpha (\log \tau / \tau)^{\alpha/2} + 2/\tau^4 \leq C (\log \tau / \tau)^{\alpha/2}. \quad (\text{B.30})$$

Taking expectations on both sides of (B.29), we obtain that

$$\begin{aligned} & \mathbb{E}[\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{E}_t, \mathcal{L}_t] \\ & \stackrel{(i)}{\leq} \left( \mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{\beta+1}} \right] + \delta \right) / \tau + 2 \mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^{1+\frac{1}{\beta+1}} \right] + 2 \mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}| \right] \delta + \mathbb{P}(\mathcal{A}_{3,t}^c | \mathcal{E}_t) + \mathbb{P}(\mathcal{A}_{4,t}^c | \mathcal{E}_t), \\ & \stackrel{(ii)}{\leq} 6 (\log \tau / \tau)^{\frac{1}{2(\beta+1)}} / \tau + \delta / \tau + 36 (\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} + 6\delta \sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c | \mathcal{E}_t) + C/\tau^2, \end{aligned} \quad (\text{B.31})$$

where inequality (i) follows from the taking expectation on both sides, and inequality (ii) follows from using (B.30) for the first, third and fourth summands, and the sixth summand follows from (B.27).

**Case (b):**  $\mathcal{E}_t^c = \{1 - B_t/\tau \leq 4\sqrt{\log \tau / \tau}\}$ . The event  $1 - B_t/\tau \leq 4\sqrt{\log \tau / \tau}$  implies that  $(\tau - B_t - 1)/\tau \leq 4\sqrt{\log \tau / \tau}$ , and obviously we have  $(q_t^l - p_t^{\text{CWG}})/q_t^l \leq 1$ . Therefore the second term in the

RHS of (B.26) is bounded above as

$$|q_t^l - p_t^{\text{CWG}}| |F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)| (\tau - B_t - 1) / (q_t^l \tau) \leq 4\sqrt{\log \tau / \tau} \left| F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+) \right|$$

Therefore we can upper bound  $\mathbb{E} \left[ \partial \mathcal{R}_t(B_t, \mu_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{E}_t^c, \mathcal{L}_t \right]$  as

$$\begin{aligned} & \mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{E}_t^c, \mathcal{L}_t \right] \\ & \leq \mathbb{1}_{A_{3,t}} \left[ \left| q_t^u - p_t^{\text{CWG}} \right|^{\frac{1}{1+\beta}} / \tau + \delta / \tau \right] + \mathbb{1}_{A_{3,t}} \left[ 4\sqrt{\log \tau / \tau} |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{\beta+1}} + 4\delta\sqrt{\log \tau / \tau} \right] + \mathbb{1}_{A_{3,t}^c}, \\ & \leq \left| q_t^u - p_t^{\text{CWG}} \right|^{\frac{1}{1+\beta}} / \tau + \delta / \tau + 4\sqrt{\log \tau / \tau} |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{\beta+1}} + 4\delta\sqrt{\log \tau / \tau} + \mathbb{1}_{A_{3,t}^c}, \end{aligned}$$

Taking expectations on both sides we get that

$$\begin{aligned} & \mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{E}_t^c, \mathcal{L}_t \right] \\ & \leq \mathbb{E} \left[ \left| q_t^u - p_t^{\text{CWG}} \right|^{\frac{1}{1+\beta}} / \tau + \delta / \tau + 4\sqrt{\log \tau / \tau} \mathbb{E} \left[ \left| q_t^u - p_t^{\text{CWG}} \right|^{\frac{1}{1+\beta}} \right] + 4\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c | \mathcal{E}_t^c), \right. \\ & \left. \leq 6(\log \tau / \tau)^{\frac{1}{2(1+\beta)}} / \tau + \delta / \tau + 24(\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} + 4\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c | \mathcal{E}_t^c), \right. \end{aligned} \quad (\text{B.32})$$

where the second inequality follows from the fact that the first and the second term are bounded by (B.30). This completes for event  $\mathcal{E}_t^c$ . From Lemma 18, it follows that  $\mathbb{P}(\mathcal{A}_{3,t}^c) \leq 2n(n+1)/\tau^4$ .

$$\sum_{t=1}^{\tilde{T}} \mathbb{P}(\mathcal{A}_{3,t}^c) \leq \sum_{\tau=\tau_0}^T 2n(n+1)/\tau^4 \leq n(n+1)/\tau_0^3 \leq n(n+1)\varepsilon_0^6 \stackrel{(\star)}{\leq} 1 \quad (\text{B.33})$$

where  $(\star)$  follows since if there are  $n$  gaps, there are  $n+1$  clusters, and hence  $\varepsilon_0 \leq 1/(n+1)$ .

Combining (B.31) and (B.32), for a constant  $C < \infty$  we have that

$$\mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{L}_t \right] \leq C(\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} + C\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c).$$

An identical bound holds for the regret contribution from the event  $\mathcal{H}_t$  where  $\mathcal{H}_t = \{p_t^{\text{CWG}} \geq q_t^u\}$ , by a symmetric argument. As a result, we can bound the expected total regret at time  $t$  as per  $\mathbb{E} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t})] \leq 2\mathbb{E} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{L}_t]$ . Therefore we have that there exists a constant  $C < \infty$  such that

$$\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}) \leq C(\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} + C\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c) \quad (\text{B.34})$$

Note that the RHS for (B.34) does not depend on the remaining budget  $B_t$  and hence we have a uniform bound on the expected compensation given below.

$$\sup_{B_t \geq 0} \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}) \leq C(\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} + C\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c) \quad (\text{B.35})$$

This further implies that

$$\mathbb{E}_{B_t} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}})] \leq C(\log \tau / \tau)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} + C\delta\sqrt{\log \tau / \tau} + \mathbb{P}(\mathcal{A}_{3,t}^c) \quad (\text{B.36})$$

Now summing this bound from  $t = 1$  to  $t = \tilde{T}$ , we have, using (B.33) that for a constant  $C < \infty$ , we have that

$$\begin{aligned} \sum_{t=1}^{\tilde{T}} \mathbb{E}_{B_t} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}})] &\leq C \left[ (1 + 1/\beta)(\log T)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}} \cdot \mathbb{1}\{\beta > 0\} + (\log T)^2 \mathbb{1}\{\beta = 0\} \right] \\ &\quad + C\delta\sqrt{T \log T}. \end{aligned}$$

Finally, consider time steps  $t$  such that  $\tilde{T} + 1 \leq t \leq T$ . In the last  $64 \log(1/\varepsilon_0)/\varepsilon_0^2$  time steps, we make use of the static allocation policy and as noted before the regret accrued during the static allocation policy is upper bounded by  $C\sqrt{\tau_0} = C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0$  for some universal constant  $C < \infty$ . Combining the two parts completes the proof.  $\blacksquare$

## B.2.6 Proof of Helper Lemmas

*Proof of Lemma 16.* Let us assume that  $p_t^{\text{CE}} \geq p_t^{\text{HTG}} + \sqrt{2 \log \tau / \tau}$ . Now conditional on  $B_t$  and given the knowledge of  $p_t^{\text{HTG}}$ , we know that there are  $B_t$  arrivals with quantile larger than  $p_t^{\text{HTG}}$  and  $\tau - B_t$  arrivals with quantiles less than  $p_t^{\text{HTG}}$ . Let  $X_t \triangleq \text{Ber}(\tau, (p_t^{\text{CE}} - \sqrt{2 \log \tau / \tau})^+)$  with  $\mathbb{E}[X_t | B_t] = (\tau - B_t - \sqrt{2 \log \tau / \tau})^+$ . Then we have that

$$\mathbb{P}\left(p_t^{\text{CE}} \geq p_t^{\text{HTG}} + \sqrt{2 \log \tau / \tau} \middle| B_t\right) \leq \mathbb{P}\left(X_t \geq \tau - B_t \middle| B_t\right) \leq \mathbb{P}\left(X_t - \mathbb{E}[X_t | B_t] \geq \sqrt{2 \log \tau / \tau} \middle| B_t\right) \leq 1/\tau^4$$

where the last inequality follows from the Hoeffding inequality. It follows that  $\mathbb{P}(p_t^{\text{CE}} \geq p_t^{\text{HTG}} + \sqrt{2 \log \tau / \tau}) \leq 1/\tau^4$ . Analogously, we can show the same for the case of  $p_t^{\text{HTG}} \geq p_t^{\text{CE}} + \sqrt{2 \log \tau / \tau}$ . ■

*Proof of Lemma 17.* We have that  $|p_t^{\text{CWG}} - p_t^{\text{HTG}}| = |p_t^{\text{CWG}} - p_t^{\text{CE}} + p_t^{\text{CE}} - p_t^{\text{HTG}}| \leq |p_t^{\text{CWG}} - p_t^{\text{CE}}| + |p_t^{\text{CE}} - p_t^{\text{HTG}}|$ . By the definition of the algorithm we have that  $|p_t^{\text{CWG}} - p_t^{\text{CE}}| \leq \sqrt{2 \log \tau / \tau}$ . Now conditional on  $B_t$ , we have that event  $\mathcal{A}_{1,t}$  implies the event  $\mathcal{A}_{2,t}$  and hence we have that  $\mathcal{A}_{2,t}^c$  implies  $\mathcal{A}_{1,t}^c$  which implies that  $\mathbb{P}(\mathcal{A}_{2,t}^c | B_t) \leq \mathbb{P}(\mathcal{A}_{1,t}^c | B_t)$ . Using the proof of Lemma 16, we have that  $\mathbb{P}(\mathcal{A}_{2,t}^c | B_t) \leq 2/\tau^4$  and the claim of the lemma follows. ■

*Proof of Lemma 18.* We have that  $\mathcal{A}_{3,t}^c = \cup_{i,j: Q_i^\circ \cap Q_j^\circ = \emptyset} \{p_t^{\text{CWG}} \in Q_i, p_t^{\text{HTG}} \in Q_j\}$  where  $A^\circ$  denotes the interior of the set  $A$ . Consider the event  $\{p_t^{\text{CWG}} \in Q_i, p_t^{\text{HTG}} \in Q_j\}$  such that  $Q_i^\circ \cap Q_j^\circ = \emptyset$ . From the definition of  $p_t^{\text{CWG}}$  in Algorithm 2 and the fact that  $Q_i^\circ \cap Q_j^\circ = \emptyset$  implies that  $|p_t^{\text{CE}} - p_t^{\text{HTG}}| \geq \sqrt{2 \log \tau / \tau}$ . Using Lemma 16 and the union bound completes the proof. ■

*Proof of Lemma 19.* Assume that  $p_t^{\text{CWG}} \leq q_t^l$ , then according to the definition of  $p_t^{\text{HTG}}$ , we have that  $p_t^{\text{HTG}} = q_t^u$ . Compensation is provided only if  $q_t^\theta \in [p_t^{\text{CWG}}, p_t^{\text{HTG}}]$ . Suppose that is the case, then we have that  $F^{-1}((p_t^{\text{CWG}})^+) \leq \theta_t \leq F^{-1}(p_t^{\text{CWG}}) = F^{-1}(q_t^u) = u_t$ . The CWG policy would accept the candidate with ability  $\theta_t$  since  $\theta_t \geq F^{-1}((p_t^{\text{CWG}})^+)$  where as the HTG would want to reject the candidate with ability  $\theta_t$ , because in the future it knows that it can select a candidate with ability at least  $u_t \geq \theta_t$ . Hence to persuade the HTG, we need to compensate it  $u_t - \theta_t = F^{-1}(p_t^{\text{HTG}}) - \theta_t$  and maximum compensation can hence be  $F^{-1}(p_t^{\text{HTG}}) - F^{-1}((p_t^{\text{CWG}})^+)$ .

An analogous analysis can be done for the case when  $p_t^{\text{CWG}} \geq q_t^u$  which follows similarly.  $\blacksquare$

### B.3 Proof of Corollary 5

*Proof of Corollary 5.* The discrete distribution considered is a  $(\beta = 0, \varepsilon_0, \delta = 0)$ -clustered distribution for  $\varepsilon_0 = \min_{1 \leq i \leq m} \{f_i\}$ . As done for the general case above, our analysis for the case of discrete distributions as considered in the Example 1 also follows in two parts. The regret accrued during the second part due to the static allocation policy is upper bounded by  $C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0$  for some universal constant  $C < \infty$ . Next we will consider the first part. The argument for the first part will mirror the analysis presented in the proof of Theorem 5 except for one important improvement we make for this special case. Consider the regret contribution of sample paths satisfying  $\mathcal{L}_t \triangleq \{p_t^{\text{CWG}} \leq q_t^l\}$  as we did previously. (Again, there is a analogous analysis for the symmetric event  $\mathcal{H}_t \triangleq \{p_t^{\text{CWG}} > q_t^u\}$ , which we omit to avoid repetition.) The only but important distinction in the case of discrete distributions is that on the event  $\mathcal{A}_{3,t}$ , which is that  $q_t^l, q_t^u$  and  $p_t^{\text{CWG}}$  are quantiles belonging to the same cluster, the compensation is given as  $F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)$ , however for discrete distributions, we have that  $F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+) = 0$ . Previously, in the general case, we had upper bounded  $F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)$  by  $|q_t^u - p_t^{\text{CWG}}|^{1/(1+\beta)} + \delta$  using Definition 2. Because  $F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+) = 0$  on the event  $\mathcal{A}_{3,t}$ , from (B.26), we have that

$$\mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t, \mathcal{E}_t \right] \leq \mathbb{1}_{\mathcal{A}_{3,t}^c} + \mathbb{1}_{\mathcal{A}_{4,t}^c},$$

because  $\sup_{B_t, a_t, \theta_{\geq t}} \partial \mathcal{R}_t(B_t, a_t; \theta_{\geq t}) \leq 1$ . Taking expectations on both sides, we have that

$$\mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{L}_t \right] \leq \mathbb{P}(\mathcal{A}_{3,t}^c) + \mathbb{P}(\mathcal{A}_{4,t}^c) \leq \mathbb{P}(\mathcal{A}_{3,t}^c) + C/\tau^2,$$

Summing this upper bound from  $\tau = \tau_0$  to  $\tau = T$ , we get that the summation is upper bounded by a universal constant  $C < \infty$  using (B.33). Combining the regret accrued in the two parts, we get the required result.  $\blacksquare$

## B.4 Proof of Corollary 6

*Proof of Corollary 6.* Since by assumption, there are no gaps in the distribution, we have that  $p_t^{\text{CWG}} = p_t^{\text{CE}}$  for all  $t$ . Our analysis will follow along the same lines as the analysis for Theorem 5 with  $\delta = 0$ . From (B.26), we have that

$$\begin{aligned} \mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t \right] &\leq \frac{|F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)|}{\tau} \\ &+ \frac{(q_t^l - p_t^{\text{CWG}}) |F^{-1}(q_t^u) - F^{-1}((p_t^{\text{CWG}})^+)| (\tau - B_t - 1)}{q_t^l \tau} \end{aligned}$$

This is where our proof departs from the proof of Theorem 5. The fact that the CWG policy boils down to the CE policy when there are no non-trivial gaps simplifies the analysis to a great extent. Instead of considering two cases to bound the ratio  $(\tau - B_t - 1)/(q_t^l \tau)$ , we can bound it much simply. From the definition of  $p_t^{\text{CE}}$ , we have that  $p_t^{\text{CE}} = 1 - B_t/\tau$ , which implies that  $(\tau - B_t - 1)/(q_t^l \tau) \leq p_t^{\text{CE}} \tau / (q_t^l \tau) \leq 1$  since  $p_t^{\text{CWG}} = p_t^{\text{CE}}$  and we are considering the sample paths on which  $p_t^{\text{CWG}} \leq q_t^l$ . Since  $F$  is a  $(\beta, \varepsilon_0 = 1, \delta = 0)$ -clustered distribution, we have that

$$\mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{F}_t, q_t^l, q_t^u, \mathcal{L}_t \right] \leq \frac{|q_t^u - p_t^{\text{CWG}}|^{\frac{1}{1+\beta}}}{\tau} + |q_t^u - p_t^{\text{CWG}}|^{1+\frac{1}{1+\beta}}$$

Taking expectations, we have that

$$\begin{aligned} \mathbb{E} \left[ \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | L_t \right] &\leq \mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^{\frac{1}{1+\beta}} \right] / \tau + \mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^{1+\frac{1}{1+\beta}} \right] \\ &\leq C \left( \tau^{-\frac{1}{2(1+\beta)}-1} + \tau^{-\frac{1}{2(1+\beta)}-\frac{1}{2}} \right), \end{aligned}$$

where the second inequality follows from the fact that  $\mathbb{E} [|q_t^u - p_t^{\text{CWG}}|^\alpha] \leq C(T - t + 1)^{-\alpha/2}$  for any  $\alpha \in (0, 2]$  and the fact that  $1/(1 + \beta) \in (0, 1]$  and  $1 + 1/(1 + \beta) \in (0, 2]$ . Recall that  $\mathbb{E} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t})] \leq 2\mathbb{E} [\partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}, \theta_{\geq t}) | \mathcal{L}_t]$ . Summing this over  $T$  time steps gives us the regret scaling in Corollary 6. To complete the proof, we will prove that for all  $t \leq T - 1$  and  $\alpha \in (0, 2]$ , we have that  $\mathbb{E} [|q_t^u - p_t^{\text{CWG}}|^\alpha] \leq C(T - t + 1)^{-\alpha/2}$ . This inequality follows from the

Hoeffding inequality as shown below.

$$\mathbb{E} \left[ |q_t^u - p_t^{\text{CWG}}|^\alpha \right] = \int_0^\infty \mathbb{P} \left( |q_t^u - p_t^{\text{CWG}}|^\alpha \geq x \right) dx \leq 2 \int_0^\infty \exp \left( -2\tau x^{2/\alpha} \right) dx = 2^{-\alpha/2} \alpha \Gamma(1 + 2/\alpha) \tau^{-\alpha/2},$$

This completes the proof of Corollary 6. ■

## B.5 Recovering existing regret guarantees for RAMS

In this section, we provide corollaries which show that RAMS attains near-optimal regret scaling for a variety of online resource allocation problems under different assumptions. As a first application of Theorem 6, we consider the multisecretary problem. For analytical simplicity, we consider a minor variant of RAMS where for the first  $\tilde{T} = T - \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$  time steps, we implement RAMS as stated in Algorithm 3 and in the final  $\lceil 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rceil$  time steps, we implement a static threshold policy as done in the case of Algorithm 2. This minor variant of RAMS inherits the guarantees established in Theorem 5.

**Corollary 12 ( $\beta$ -dependent regret for multisecretary)** *Consider the multisecretary problem with the candidate ability distribution  $F$  being  $(\beta, \varepsilon_0, \delta)$ -clustered for some fixed  $\beta \in [0, \infty)$ ,  $\varepsilon_0 \in (0, 1]$  and  $\delta \in [0, 1]$ . Fix the parameter  $\eta > 2$  in Theorem 6. Assume that the number of sample paths drawn at time  $t$  is sufficiently large, specifically  $K_t \geq (T - t + 1)^{\eta+\nu}$  for some  $\nu > 0$ . Then there exists a constant  $C \equiv C(F, \eta, \nu) < \infty$ , such that for all  $T \in \mathbb{N}$  and  $B \in \mathbb{N}$ , the regret for RAMS is bounded above as*

$$\begin{aligned} \text{Regret}(B, T; \text{RAMS}) &\leq C(1 + 1/\beta)(\log T)^{\frac{1}{2} + \frac{1}{2(\beta+1)}} T^{\frac{1}{2} - \frac{1}{2(\beta+1)}} \cdot \mathbb{1}\{\beta > 0\} + C(\log T)^2 \mathbb{1}\{\beta = 0\} \\ &\quad + C\delta\sqrt{T \log T} + C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0. \end{aligned}$$

Next we zoom out from the multisecretary problem and consider the more general network revenue management and online matching problems. We present four assumptions under which these problems have been studied. These assumptions are stated in the notation introduced in this

paper.

**Assumption 4 (Small number of types for NRM)** *The type distribution  $F$  is supported on a discrete set  $\{(r_1, \mathbf{c}_1), (r_2, \mathbf{c}_2), \dots, (r_n, \mathbf{c}_n)\}$  with  $c_{\theta,k} \in \{0, 1\}$  for all  $\theta \in \{1, \dots, n\}, k \in \{1, \dots, d\}$ .*

**Assumption 5 (Infinitely many types for NRM with density bounded below)** *The consumption random vector  $\mathbf{c}_\theta$  is bounded i.e.  $\underline{\nu} \leq \|\mathbf{c}_\theta\|_\infty \leq \bar{\nu}$  for  $0 < \underline{\nu} \leq \bar{\nu} < \infty$  for all  $\theta \in \Theta$ . Conditional on the consumption vector  $\mathbf{c}_\theta$ , the reward distribution  $F_\theta$  is assumed to be  $(\beta = 0, \varepsilon_0 = 1)$ -clustered with reward random variable  $r_\theta$  being bounded in  $[0, 1]$ .*

**Assumption 6 (Infinitely many types for NRM)** *The consumption random vector  $\mathbf{c}_\theta$  is supported on a small discrete set  $\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$  with  $c_{\theta,k} \in \{0, 1\}$  for all  $\theta \in \{1, \dots, n\}$  and  $k \in \{1, \dots, d\}$ . Conditional on the consumption vector  $\mathbf{c}_\theta$ , the reward distribution  $F_\theta$  is assumed to be  $(\beta = 0, \varepsilon_0)$ -clustered distribution with  $\varepsilon_0 \in (0, 1]$  and the reward random variable  $r_\theta$  being bounded in  $[0, 1]$ .*

**Assumption 7 (Small number of types for Online Matching)** *The type distribution  $F$  is supported on a discrete set of reward vectors  $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$  where  $\mathbf{r}_\theta \in [0, 1]^d$  for all  $\theta \in \{1, \dots, n\}$ .*

*Discussion of the assumptions.* Recall Assumptions 2 (a few discrete types) and 3 (continuous types) for the multisecretary problem. Assumptions 4 and 7 are a natural generalization of Assumption 2 in the context of network revenue management and online matching respectively and is often a standard assumption in this literature [36, 37, 163]. Assumptions 5 and 6 are a generalization of Assumption 3 for the NRM problem with multiple resources. Assumption 5 resembles the assumption studied in [70], however Assumption 5 is stronger than the one in [70] in the sense that [70] allows for arbitrarily small consumption vectors (i.e.,  $\underline{\nu} = 0$ ) while Assumption 5 assumes that consumption vectors are bounded below. Additionally [70] allows for unbounded rewards while Assumption 5 assumes that the rewards are bounded in the interval  $[0, 1]$ . Note that we study a stronger version of the assumptions in [70] for the sake of technical simplicity and conjecture that RAMS will achieve the same logarithmic regret scaling under the assumptions studied in [70]. The key similarity between Assumption 5 and the assumption studied in [70] is that both assumptions

imply that the fluid problem is non-degenerate which enables the logarithmic regret scaling. Assumption 6, while being similar to Assumption 5, allows for degeneracy in the fluid problem and was recently studied by [87]. There are two key distinctions between Assumptions 5 and 6: (i) Assumption 6 only permits a few consumption types and (ii) Assumption 6 allows for gaps in the (conditional) reward distributions which in turn permits degeneracy in the fluid problem.

Theorem 6 tells us that RAMS inherits the regret guarantees previously established for other algorithms, under Assumptions 4-7. This is formalized in the following corollaries. Note that in all our regret guarantees provided below, the only scaling parameters are the time horizon  $T$  and the budget  $B$  and all other parameters are considered constant. Moreover, we emphasize that the distribution  $F$  is initially fixed and its parameters do not scale with the scaling parameter  $T$  and  $B$ . Therefore, the minimum probability parameter  $\varepsilon_0$  for the distributions considered in Assumptions 5 and 6 is also fixed and subsumed in the constants presented below.

**Corollary 13 (Regret for NRM)** *Consider the network revenue management problem with request distribution  $F$ . Fix the parameter  $\eta > 2$  in Theorem 6. Assume that the number of sample paths drawn at time  $t$  is large enough as per  $K_t \geq (T - t + 1)^{\eta+\nu}$  for some  $\nu > 0$ . Then there exists a constant  $C \equiv C(F, \eta, \nu) < \infty$ , such that for all  $T \in \mathbb{N}$  and  $B \in \mathbb{R}^d$ , we have that*

- (a) (Constant Regret with few types) *If  $F$  satisfies Assumption 4,  $\text{Regret}(B, T; \text{RAMS}) \leq C$ .*
- (b) (Logarithmic Regret) *If  $F$  satisfies Assumption 5,  $\text{Regret}(B, T; \text{RAMS}) \leq C \log T$ .*
- (c) (Log-Squared Regret) *If  $F$  satisfies Assumption 6,  $\text{Regret}(B, T; \text{RAMS}) \leq C \log^2 T$ .*

**Corollary 14 (Constant Regret for Online Matching)** *Consider the online matching setting with request distribution  $F$  satisfying Assumption 7. Fix the parameter  $\eta > 2$  in Theorem 6. Assume that the number of sample paths drawn at time  $t$  is large enough as per  $K_t \geq (T - t + 1)^{\eta+\nu}$  for some  $\nu > 0$ . Then there exists a constant  $C \equiv C(F, \eta, \nu) < \infty$  such that for all  $T \in \mathbb{N}$  and  $B \in \mathbb{R}^d$ , the regret for RAMS is bounded above as  $\text{Regret}(B, T; \text{RAMS}) \leq C$ .*

## B.6 Proofs Related to RAMS

### B.6.1 Proof of Claim 1

*Proof of Claim 1.* Given any budget  $B_t \geq 0$  and any sample path  $\theta_{\geq t+1}$ , if the hindsight to go (HTG) policy decides to accept the request  $\theta_t$ , we can make it reject the request  $\theta_t$  by paying a maximum compensation of  $r_{\max}$ . On the flip side, the hindsight to go policy can extract at most  $r_{\max} \bar{v}/\underline{\gamma}$  in the future for every resource  $\theta_t$  makes use of, hence if the hindsight to go (HTG) policy wants to reject  $\theta_t$ , we can make it accept the request  $\theta_t$  by paying a compensation of  $dr_{\max} \bar{v}/\underline{\gamma}$  since the request  $\theta_t$  can make use of at most  $d$  resources. ■

### B.6.2 Proof of Lemma 2

*Proof of Lemma 2.* Using (2.5), for a simulated sample path  $\theta_{\geq t}^{(i)} \triangleq \{\theta_t, \theta_{\geq t+1}^{(i)}\}$ , we have that

$$\begin{aligned} \partial \mathcal{R}_t(B_t, a, \theta_{\geq t}^{(i)}) &= V_t^{\text{hs}}(B_t; \theta_{\geq t}^{(i)}) - \left[ V_{t+1}^{\text{hs}}(B_t - c(\theta_t, a); \theta_{\geq t+1}^{(i)}) + r(\theta_t, a) \right] \\ &= V_t^{\text{hs}}(B_t; \theta_{\geq t}^{(i)}) - Q_t^{\text{hs}}(B_t, a; \theta_{\geq t}^{(i)}). \end{aligned}$$

Note that the term  $V_t^{\text{hs}}(B_t; \theta_{\geq t}^{(i)})$  does not depend on the action  $a \in \mathcal{A}(B_t, \theta_t)$  and hence we have

$$\arg \max_{a \in \mathcal{A}(B_t, \theta_t)} K_t^{-1} \sum_{i=1}^{K_t} Q_t^{\text{hs}}(B_t, a; \tilde{\theta}_{\geq t}^{(i)}) = \arg \min_{a \in \mathcal{A}(B_t, \theta_t)} K_t^{-1} \sum_{i=1}^{K_t} \partial \mathcal{R}_t(B_t, a; \tilde{\theta}_{\geq t}^{(i)}), \quad (\text{B.37})$$

i.e., RAMS takes an action  $a \in \mathcal{A}(B_t, \theta_t)$  which minimizes the simulation-based estimate of the expected marginal compensation. ■

### B.6.3 Proof of Theorem 6

*Proof of Theorem 6.* Given a budget  $B_t$  and a request  $\theta_t$  at time  $t$ , from Algorithm 3 it follows

that the action under the RAMS policy is given by:

$$a_t^{\text{RAMS}} = \arg \max_{a \in \mathcal{A}} \frac{1}{K_t} \sum_{i=1}^{K_t} Q_t^{\text{hs}} \left( B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right), \quad (\text{B.38})$$

where  $K_t$  denotes the number of simulated sample paths used at time  $t$ ,  $Q_t^{\text{hs}} \left( B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right)$  is defined in (2.4),  $\tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \triangleq \{\theta_t, \tilde{\theta}_{t+1}^{(i)}, \tilde{\theta}_{t+2}^{(i)}, \dots, \tilde{\theta}_T^{(i)}\}$  and  $\{\tilde{\theta}_{t+1}^{(i)}, \tilde{\theta}_{t+2}^{(i)}, \dots, \tilde{\theta}_T^{(i)}\}$  denote the  $i$ -th sequence of simulated sample paths. From Lemma 2, it follows that the action under the RAMS policy can be equivalently written as

$$a_t^{\text{RAMS}} = \arg \min_{a \in \mathcal{A}} \frac{1}{K_t} \sum_{i=1}^{K_t} \partial \mathcal{R}_t \left( B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right), \quad (\text{B.39})$$

where  $\partial \mathcal{R}_t \left( B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right) = \max_{\hat{a} \in \mathcal{A}} Q_t^{\text{hs}} \left( B_t, \hat{a}; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right) - Q_t^{\text{hs}} \left( B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)} \right)$ .

From the regret decomposition lemma of [36], it follows that

$$\text{Regret}(B, T; \text{RAMS}) = \sum_{t=1}^T \mathbb{E}_{B_t^{\text{RAMS}}} \left[ \partial \mathcal{R}_t \left( B_t^{\text{RAMS}}, a_t^{\text{RAMS}} \right) \right]. \quad (\text{B.40})$$

We note that  $\mathbb{E}_{B_t^{\text{RAMS}}} \left[ \partial \mathcal{R}_t \left( B_t^{\text{RAMS}}, a_t^{\text{RAMS}} \right) \right] \leq \sup_{B_t \geq 0} \partial \mathcal{R}_t(B_t, a_t^{\text{RAMS}})$  for all  $t \in \{1, 2, \dots, T\}$ .

Now to prove to Theorem 6, it suffices for us to show that

$$\sup_{B_t \geq 0} \partial \mathcal{R}_t \left( B_t, a_t^{\text{RAMS}} \right) \leq \Delta_t(\text{ALG}) + C(\eta, |\mathcal{A}|, C) K_t^{-\frac{1}{\eta}},$$

where  $\Delta_t(\text{ALG})$  is the uniform upper bound assumed in condition (i) at time  $t$  under ALG. We will begin by upper bounding the quantity  $\partial \mathcal{R}_t(B_t, a_t^{\text{RAMS}})$ . For some fixed parameter  $\eta > 2$ , conditional on the budget  $B_t$ , and request  $\theta_t$ , define the following ‘‘good’’ event  $\mathcal{G}_t$

$$\mathcal{G}_t = \cap_{a \in \mathcal{A}} \left\{ \left| K_t^{-1} \sum_{i=1}^{K_t} \partial \mathcal{R}_t(B_t, a; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)}) - \mathbb{E} [\partial \mathcal{R}_t(B_t, a; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t] \right| \leq K_t^{-\frac{1}{\eta}} \right\}. \quad (\text{B.41})$$

Using the definition of  $\partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}})$  and the tower property we have,

$$\partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}) = \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) | B_t \right] = \mathbb{E} \left[ \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right] | B_t \right]. \quad (\text{B.42})$$

Now we further write the inner (conditional) expectation  $\mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right]$  as

$$\mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right] = \underbrace{\mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) \mathbb{1}_{\mathcal{G}_t} | \theta_t, B_t \right]}_{(\spadesuit)} + \underbrace{\mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) \mathbb{1}_{\mathcal{G}_t^c} | \theta_t, B_t \right]}_{(\clubsuit)}.$$

Now we have two terms  $(\spadesuit)$  and  $(\clubsuit)$  to bound. We begin by bounding the term  $(\spadesuit)$ . We have that

$$\begin{aligned} \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) \mathbb{1}_{\mathcal{G}_t} | \theta_t, B_t \right] &\stackrel{(a)}{\leq} \left( K_t^{-1} \sum_{i=1}^{K_t} \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)}) + K_t^{-\frac{1}{\eta}} \right) \mathbb{1}_{\mathcal{G}_t}, \\ &\stackrel{(b)}{\leq} \left( K_t^{-1} \sum_{i=1}^{K_t} \partial\mathcal{R}_t(B_t, a_t^{\text{ALG}}; \tilde{\boldsymbol{\theta}}_{\geq t}^{(i)}) + K_t^{-\frac{1}{\eta}} \right) \mathbb{1}_{\mathcal{G}_t} \\ &\stackrel{(c)}{\leq} \left( \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{ALG}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right] + 2K_t^{-\frac{1}{\eta}} \right) \mathbb{1}_{\mathcal{G}_t} \\ &\stackrel{(d)}{\leq} \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{ALG}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right] + 2K_t^{-\frac{1}{\eta}} \end{aligned}$$

where (a) follows from definition of event  $\mathcal{G}_t$  applied to the action  $a_t^{\text{RAMS}}$ , (b) follows from the fact that RAMS takes the action according to (B.39), (c) follows from definition of event  $\mathcal{G}_t$  applied to the action  $a_t^{\text{ALG}}$  and (d) follows from that fact that  $\mathbb{1}_{\mathcal{G}_t} \leq 1$ . Using this it follows that

$$(\spadesuit) = \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \boldsymbol{\theta}_{\geq t}) \mathbb{1}_{\mathcal{G}_t} | \theta_t, B_t \right] \leq \mathbb{E} \left[ \partial\mathcal{R}_t(B_t, a_t^{\text{ALG}}; \boldsymbol{\theta}_{\geq t}) | \theta_t, B_t \right] + 2K_t^{-\frac{1}{\eta}}$$

Next we bound the term  $(\clubsuit)$ . Define  $Y_i(a) \triangleq \partial\mathcal{R}_t(B_t, a; \boldsymbol{\theta}_{\geq t}^{(i)})$ . From Assumption (ii) in Theorem 6, we have that  $\partial\mathcal{R}_t(B_t, a; \boldsymbol{\theta}_{\geq t}) \leq C$  almost surely for all  $B_t \geq 0, a \in \mathcal{A}$  and  $\boldsymbol{\theta}_{\geq t}$ . Therefore we have that  $\{Y_i(a)\}_{i=1}^{K_t}$  are i.i.d random variables with  $|Y_i(a)| \leq C$  almost surely. Hence we have

that

$$(\clubsuit) \leq C\mathbb{E}[\mathbb{1}_{\mathcal{G}_t^c} | \theta_t, B_t] = C\mathbb{P}(\mathcal{G}_t^c | \theta_t, B_t) \leq C(\eta, |\mathcal{A}|, C)K_t^{-\frac{1}{\eta}}$$

where the last inequality follows from union bound and Hoeffding's inequality as described below.

$$\begin{aligned} \mathbb{P}(\mathcal{G}_t^c | \theta_t, B_t) &\stackrel{(a)}{\leq} \sum_{a \in \mathcal{A}} \mathbb{P}\left(\left|K_t^{-1} \sum_{i=1}^{K_t} Y_i(a) - \mathbb{E}[Y_i(a) | \theta_t, B_t]\right| > K_t^{-\frac{1}{\eta}}\right), \\ &\stackrel{(b)}{\leq} 2|\mathcal{A}| \exp\left(-\frac{2K_t^{2-\frac{2}{\eta}}}{K_t C^2}\right), \\ &\stackrel{(c)}{\leq} 2|\mathcal{A}| \exp\left(2K_t^{\frac{\eta-2}{\eta}} / C^2\right), \\ &\stackrel{(d)}{\leq} C(\eta, |\mathcal{A}|, C)K_t^{-\frac{1}{\eta}}, \end{aligned}$$

where (a) follows from union bound, (b) follows from Hoeffding's inequality, (c) follows trivially, (d) follows for some appropriate constant  $C(\eta, |\mathcal{A}|, C)$  since  $\exp(-x) \leq C(p)x^{-p}$  for some  $p > 0$ .

Given the bound on  $(\spadesuit)$  and  $(\clubsuit)$ , we have that

$$\mathbb{E}\left[\partial \mathcal{R}_t(B_t, a_t^{\text{RAMS}}; \theta_{\geq t}) | \theta_t, B_t\right] \leq \mathbb{E}\left[\partial \mathcal{R}_t(B_t, a_t^{\text{ALG}}; \theta_{\geq t}) | \theta_t, B_t\right] + C(\eta, |\mathcal{A}|, C)K_t^{-\frac{1}{\eta}}.$$

Using (B.42), we have that

$$\partial \mathcal{R}_t(B_t, a_t^{\text{RAMS}}) \leq \partial \mathcal{R}_t(B_t, a_t^{\text{ALG}}) + C(\eta, |\mathcal{A}|, C)K_t^{-\frac{1}{\eta}}.$$

Taking a supremum over the budget  $B_t \geq \mathbf{0}$ , we have that

$$\sup_{B_t \geq \mathbf{0}} \partial \mathcal{R}_t(B_t, a_t^{\text{RAMS}}) \leq \sup_{B_t \geq \mathbf{0}} \partial \mathcal{R}_t(B_t, a_t^{\text{ALG}}) + C(\eta, |\mathcal{A}|, C(F))K_t^{-\frac{1}{\eta}} \leq \Delta_t(\text{ALG}) + C(\eta, |\mathcal{A}|, C(F))K_t^{-\frac{1}{\eta}},$$

where the last inequality follows from Assumption (i). This completes the proof.  $\blacksquare$

### B.6.4 Proof of Corollaries 12, 13 and 14

From Theorem 6, recall that the regret upper bound for RAMS (or minor variants of RAMS) can be decomposed as a sum of the following two terms

$$\text{Regret}(B, T; \text{RAMS}) \leq \underbrace{\sum_{t=1}^T \Delta_t(\text{ALG})}_{(\diamond)} + C \underbrace{\sum_{t=1}^T K_t^{-\frac{1}{\eta}}}_{(\heartsuit)}$$

where the constant  $C < \infty$  is a function of the parameter  $\eta$ , size of the action set  $|\mathcal{A}|$  and the distribution  $F$ . Note that  $(\heartsuit)$  is common across different problem settings and assumptions while  $(\diamond)$  needs to be dealt with separately. For each Corollary 12, 13 and 14, we have that  $K_t \geq (T - t + 1)^{\eta + \nu}$  where  $\eta > 2$  is a fixed parameter from Theorem 6 and  $\nu > 0$  is a chosen parameter. This implies that  $K_t^{-\frac{1}{\eta}} \leq (T - t)^{-1 - \frac{\nu}{\eta}}$  and hence we have that  $C \sum_{t=1}^T K_t^{-\frac{1}{\eta}} \leq C \sum_{t=1}^T (T - t + 1)^{-1 - \frac{\nu}{\eta}} \leq C \int_1^T x^{-1 - \frac{\nu}{\eta}} dx \leq C(\eta, |\mathcal{A}|, F, \nu)$  since  $\nu/\eta > 0$ . Since this is common across all the corollaries, we have that the contribution to regret due to the number of simulated sample paths  $K_t$  is a constant (depending on  $\eta, \nu, F$  and  $|\mathcal{A}|$ ). The only thing remaining to bound is  $(\diamond)$  under different assumptions and problem settings.

*Proof of Corollary 12.* From Lemma 15, it follows that for  $t \leq \tilde{T} = T - \lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$ , we have that  $\sup_{B_{t \geq 0}} \partial \mathcal{R}_t(B_t, a_t^{\text{CWG}}) \leq C \left( (\log \tau/\tau)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} + \delta \sqrt{\log \tau/\tau} \right)$  which implies that for  $t \leq \tilde{T}$ ,  $\Delta_t(\text{CWG}) = C \left( (\log \tau/\tau)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} + \delta \sqrt{\log \tau/\tau} \right)$ . Summing  $\Delta_t(\text{CWG})$  from  $t = 1$  to  $t = \tilde{T}$ , implies that the regret contribution is at most  $C((\log T)^{\frac{1}{2} + \frac{1}{2(1+\beta)}} T^{\frac{1}{2} - \frac{1}{2(1+\beta)}} \mathbb{1}\{\beta > 0\} + \log^2 T \mathbb{1}\{\beta = 0\} + \delta \sqrt{T \log T})$ . Since we are considering a variant of RAMS which employs a static allocation policy (same as the one deployed in Algorithm 2) for the last  $\lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$ , the regret accrued over the last  $\lfloor 64 \log(1/\varepsilon_0)/\varepsilon_0^2 \rfloor$  time steps is upper bounded as  $C\sqrt{\log(1/\varepsilon_0)}/\varepsilon_0$ . Adding up all the contributions (including due to  $(\heartsuit)$ ), we attain the same regret scaling as in Theorem 5. ■

*Proof of Corollary 13.* For each of the Assumptions 4, 5 and 6, we have that  $\sup_{B \geq 0, a \in \mathcal{A}, \theta_{\geq t}} \partial \mathcal{R}_t(B, a; \theta_{\geq t}) \leq d$  for all  $t \in \{1, \dots, T\}$  since the offline will need a compensation of atmost  $r_{\max} = 1$  per resource in the future for accepting or rejecting the request  $\theta_t$ . Since there are  $d$  fixed resources, the compen-

sation is atmost  $d$ . Now under different assumptions, we have different algorithms with different values for  $\Delta_t(\text{ALG})$ .

- (a) Under Assumption 4. From (9) in [36], we have that for the Bayes Selector algorithm described in Algorithm 2 of [36],  $\sup_{B_t \geq 0} \partial \mathcal{R}_t(B_t, a_t^{\text{BayesSelector}}) \leq d \exp(-c\tau) := \Delta_t(\text{BayesSelector})$  for  $t \leq T - T_0$  where  $\tau = T - t + 1$ , and  $c, T_0$  are constants which depend only on the distribution  $F$ . Using the fact that in the last constant  $T_0$ , the regret accrued is atmost  $dT_0$  and  $\int_1^T d \exp(-c\tau) d\tau \leq C$ , we have that the total regret accrued by RAMS under Assumption 4 is at most a constant  $C$  which depends on the parameters  $\eta > 2, \nu > 0$ , number of resources  $d$  and the distribution  $F$ .
- (b) Under Assumption 5. From Lemma 5, 8, 9 and 10 of [87], for the Bid Price algorithm described in Algorithm 3 of [87], we have that  $\sup_{B_t \geq 0} \partial \mathcal{R}_t(B_t, a_t^{\text{BidPrice}}) \leq C/\tau := \Delta_t(\text{BidPrice})$  for  $t \leq T - T_0$  where  $\tau = T - t + 1$  and  $C, T_0$  are constants which depend only on the distribution  $F$ . Using the fact that in the last constant  $T_0$ , the regret accrued is atmost  $dT_0$  and  $\int_1^T C/\tau d\tau \leq C \log T$ , we have that the total regret accrued by RAMS under Assumption 5 is at most  $C \log T$  where the constant depends on the parameters  $\eta > 2, \nu > 0$ , number of resources  $d$  and the distribution  $F$ .
- (c) Under Assumption 6. From Theorem 1 of [87], for the Boundary Attracted algorithm described in Algorithm 2 of [87], we have that  $\sup_{B_t \geq 0} \partial \mathcal{R}_t(B_t, a_t^{\text{BoundaryAttracted}}) \leq C \log \tau/\tau := \Delta_t(\text{BoundaryAttracted})$  for  $t \leq T - T_0$  where  $\tau = T - t + 1$  and  $C, T_0$  are constants which depend only on the distribution  $F$ . Using the fact that in the last constant  $T_0$ , the regret accrued is atmost  $dT_0$  and  $\int_1^T C \log \tau/\tau d\tau \leq C \log^2 T$ , we have that the total regret accrued by RAMS under Assumption 6 is at most  $C \log^2 T$  where the constant which depends on the parameters  $\eta > 2, \nu > 0$ , number of resources  $d$  and the distribution  $F$ .

This concludes the proof for all three cases. ■

*Proof of Corollary 14.* The proof follows analogously to the proof of Corollary 13 under Assumption 4. ■

## B.7 Relating the order fulfillment problem to the multisecretary problem

### B.7.1 Motivating Example

Let's explore the following example to illuminate our point: Consider two Amazon fulfillment centers, located respectively in Salt Lake City, Utah, and Sacramento, California, as presented in Figure B.2. Both states have a total of over two thousand zip codes, which are spatially clustered and represent distinct demand locations.

The United States sees an estimated total demand volume of around hundred million Amazon packages delivered weekly [91]. Without a precise state-wise breakdown of these deliveries, we can reasonably assume that combined deliveries in California and Utah amount to no more than five million each week. Based on these figures, we calculate a total demand volume ( $T$ ) of  $5 \times 10^6$ , and a total number of demand locations or types ( $D$ ) of  $2 \times 10^3$ .

Assuming uniform demand across these locations, our model suggests that at any given time  $t$ , the probability of receiving a demand request from type  $j$  is  $D^{-1} \approx T^{-\frac{1}{2}}$ , which scales with the total demand volume. This differs from settings studied previously, which considered atomic distributions with a few types and implicitly assumed that the probability of receiving a demand request at a given time was independent of the total demand volume - an assumption inconsistent with the example we have described.

Alternatively, we could consider the setting where infinitely many types exist over a contiguous support. However, this approximation falls short in the presence of natural geographical features like the Sierra Nevada desert which creates gaps, as depicted in Figure B.2. Neither of these previously explored models satisfactorily fit this stylized order fulfillment problem. Instead, what we encounter is a scenario characterized by many types with gaps. demand request at time  $t$  is independent of the total demand volume and hence does not align well with the aforementioned example. On the other extreme, one could consider the setting with infinitely many types over a contiguous support but clearly such an approximation is wanting in the presence of gaps introduced by natural geographical features like the desert in Nevada as shown in Figure B.2. Therefore neither of the

previously studied models are a good fit for this stylized order fulfillment problem. What we have are essentially *many types with gaps*.

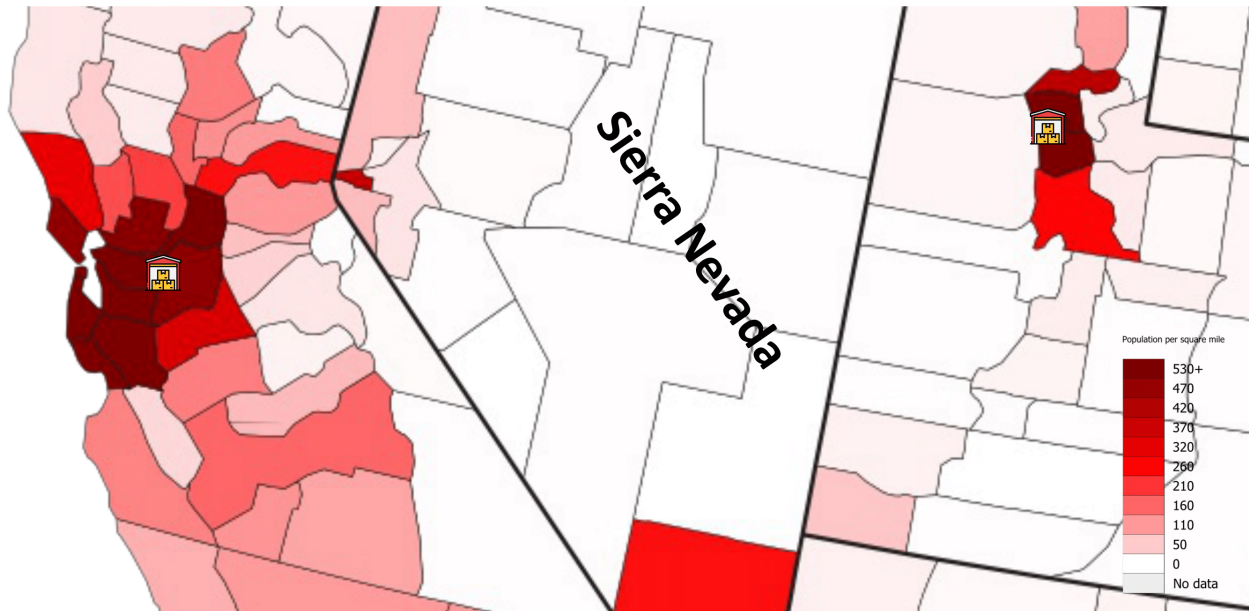


Figure B.2: Illustration of spatially distributed demand with two fulfillment centers for the order fulfillment problem

### B.7.2 Stylized model of order fulfillment

Inspired by our example illustrated above, we consider a stylized order fulfillment problem with the demand locations being spatially distributed over the unit square  $[0, 1]^2$  and two fulfillment centers (FCs) denoted as  $FCA$  and  $FCB$  with a total inventory in the two warehouses being  $T$ . The initial inventory in  $FCA$  and  $FCB$  is denoted as  $I_1^A$  and  $I_1^B$  respectively. Now at each time  $t$ , a request  $\xi_t$  arrives given by the coordinates  $(x_t, y_t) \in [0, 1]^2$  which is drawn from some spatial demand distribution  $Q$  with measure  $\mu_Q$ . Given the inventory levels  $I_t^A$  and  $I_t^B$  at time  $t$ , the order fulfillment problem is to decide which fulfillment center to serve the request  $\xi_t$  from. The goal is to minimize the total matching distance between the requests and the fulfillment center from which they are served. It is easy to see that this problem can be easily translated into the multisecretary problem. We will illustrate this correspond via an example as shown in Figure B.3.

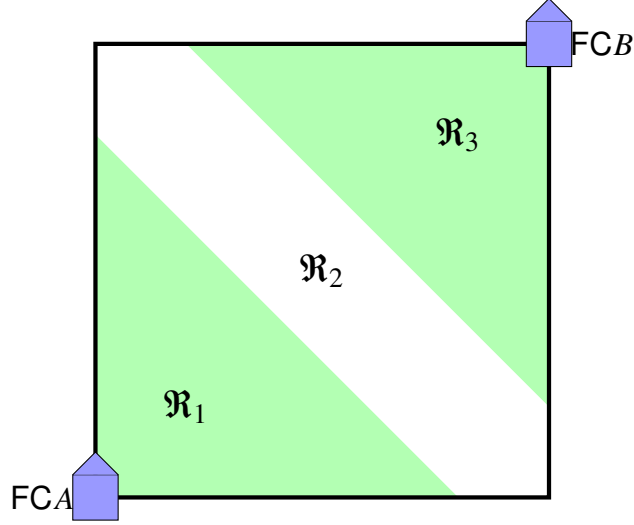


Figure B.3: Stylized example for order fulfillment with no demand from region  $\mathfrak{R}_2$

In the stylized example illustrated in Figure B.3, we assume that the demand locations are uniformly distributed in regions  $\mathfrak{R}_1$  and  $\mathfrak{R}_3$  with no demand in region  $\mathfrak{R}_2$ . The two fulfillment centers FCA and FCB are located at  $(0, 0)$  and  $(1, 1)$  respectively. Let  $d_A((x, y)) = |x| + |y|$  and  $d_B((x, y)) = |1 - x| + |1 - y|$  denote the (Manhattan) distance from the fulfillment centers FCA and FCB respectively. The hindsight optimal problem for the order fulfillment problem is the following integer program.

$$\begin{aligned}
 \min_{z_t} \quad & \sum_{t=1}^T d_A((x_t, y_t))z_t + d_B((x_t, y_t))(1 - z_t) \\
 \text{s.t.} \quad & \sum_{t=1}^T z_t = I_1^A \\
 & z_t \in \{0, 1\}, \quad \forall t
 \end{aligned} \tag{B.43}$$

The objective in (B.43) can be equivalently written as  $\sum_{t=1}^T (d_A((x_t, y_t)) - d_B((x_t, y_t))) z_t + d_B((x_t, y_t))$

and hence we can cast the minimization problem into the following maximization problem.

$$\begin{aligned}
\max_{z_t} \quad & \sum_{t=1}^T (d_B((x_t, y_t)) - d_A((x_t, y_t))) z_t \\
\text{s.t.} \quad & \sum_{t=1}^T z_t = I_1^A \\
& z_t \in \{0, 1\}, \quad \forall t
\end{aligned} \tag{B.44}$$

The optimization problem in (B.44) is the multisecretary problem with reward  $\tilde{r}((x_t, y_t)) = d_B((x_t, y_t)) - d_A((x_t, y_t))$  we consider in this work with appropriate scaling. Observe that  $\tilde{r}((x_t, y_t)) \in [-2, 2]$ , therefore we can scale the reward and define the types as  $\theta_t = (\tilde{r}((x_t, y_t)) + 2)/4 \in [0, 1]$ . We can translate the spatial demand distribution  $Q$  into the distribution over the types  $\theta_t$  as follows.

$$\begin{aligned}
\mathbb{P}(\theta_t \leq z) &= \mathbb{P}(\tilde{r}((x, y)) \leq 4z - 2) \\
&= \mathbb{P}(d_B((x, y)) - d_A((x, y)) \leq 4z - 2) \\
&= \mathbb{P}((1 - x + 1 - y) - (x + y) \leq 4z - 2) \\
&= \mathbb{P}(x + y \geq 2 - 2z)
\end{aligned}$$

Using the fact that demand locations are uniformly distribution in regions  $\mathfrak{R}_1$  and  $\mathfrak{R}_3$ , we have that

$$\mathbb{P}(\theta_t \leq z) = \begin{cases} \frac{25}{8}z^2 & z \in [0, \frac{2}{5}], \\ \frac{1}{2} & z \in [\frac{2}{5}, \frac{3}{5}], \\ 1 - \frac{25}{8}(1 - z)^2 & z \in [\frac{3}{5}, 1]. \end{cases}$$

Note that this is a  $(\beta = 0, \varepsilon_0 = \frac{1}{2})$ -clustered distribution with a gap interval  $[\frac{2}{5}, \frac{3}{5}]$ . Note that the gap in the demand location for the order fulfillment translates into a gap in the type distribution for the multisecretary problem. Moreover for simplicity we assume that the demand locations in regions  $\mathfrak{R}_1$  and  $\mathfrak{R}_3$  are distributed over a contiguous support but we can further discretize these

regions into many small types which are clustered close to each other. This captures the more realistic setting where various zipcodes are spatially close to each other. In the context of the multisecretary problem, this is captured via the  $(\beta, \varepsilon_0, \delta)$ -clustered distribution (recall Definition 2).

## B.8 Approximation of a distribution by $(\beta, \varepsilon_0, \delta)$ -clustered distributions

In this section, we illustrate that even distributions that fall outside of the class of  $(\beta, \varepsilon_0, \delta)$ -clustered distributions can be approximated by a member of this class. To so, we focus a classical singular distribution (i.e., a distribution that is not absolutely continuous and not discrete): the Cantor distribution (cf. [chung2001course]). Fix a  $n \in \mathbb{N}$  and set  $\delta = 3^{-n}$ . All points in the support of the Cantor distribution (known as the Cantor set) that are at most  $\delta$  apart are considered to be a part of a single mass cluster. Now there are  $2^n$  mass clusters given as  $[0, 3^{-n}] \cup [2 \cdot 3^{-n}, 3^{-n+1}] \cup \dots \cup [1 - 3^{-n}, 1]$  and the probability density is uniform in each of the mass clusters. This approximation of the Cantor distribution results in a  $(\beta = 0, \varepsilon_0 = 2^{-n}, \delta = 3^{-n})$ -clustered distribution, where the choice of  $n$  determines the quality of the approximation.

## B.9 Representation through $(\beta, \varepsilon_0, \delta)$ -clustered distributions

As mentioned in Section 2.3.1, there is some flexibility in how we may model a distribution or define clusters. Additionally, the parameter  $\delta$  allows us to model distributions with many *small* types. In this section, we will discuss how the same distribution can have different characterizations due to different clustering and choice of parameters  $\beta, \varepsilon_0$  and  $\delta$ . We will discuss this using two examples. For each of the examples, we will discuss two different possible clusterings and their impact on the regret guarantees. The two examples we will consider will be atomic distributions

and let  $F$  be the continuous limit of those atomic distributions described below.

$$F(x) = \begin{cases} -8(1/4 - x)^2 + 1/2, & 0 \leq x \leq 1/4 \\ 1/2, & 1/4 \leq x \leq 3/4 \\ 8(x - 3/4)^2 + 1/2, & 3/4 \leq x \leq 1. \end{cases}$$

Note that it can be easily verified that the distribution  $F$  above is a  $(\beta = 1, \varepsilon_0 = 1/2, \delta = 0)$ -clustered distribution. Let us consider two other atomic distributions with probability mass functions denoted as  $p_1$  and  $p_2$  respectively and defined using the parameters  $\eta_1$  and  $\eta_2$  as follows,

$$p_1(1/4 - k\eta_1) = p_1(3/4 + k\eta_1) = 16\eta_1^2/(4\eta_1 + 1), \forall k \in \{0, 1, \dots, 1/4\eta_1\}$$

$$p_2(1/4 - k\eta_2) = p_2(3/4 + k\eta_2) = 16\eta_2^2/(4\eta_2 + 1), \forall k \in \{0, 1, \dots, 1/4\eta_2\}$$

For  $\eta_1 = 1/24$ , we get that the distribution with probability mass function  $p_1$  is supported on twelve points and is an example of distribution with a few types (refer to center figure in Figure B.4) with the minimum probability mass being  $1/42$ . For  $\eta_2 = 1/2400$ , we get the distribution with probability mass function  $p_2$  is supported on twelve hundred points and can be considered an example of many small points since the number of types are large (1200) and each type has a small probability mass (at most  $2 \times 10^{-3}$ ) (refer to the right figure in Figure B.4). Note that as  $\eta_1, \eta_2 \rightarrow 0$ , we have that  $p_1, p_2 \rightarrow f$ .

There are two natural ways that the distribution  $p_1$  can be modelled as a  $(\beta, \varepsilon_0, \delta)$ -clustered distribution. First way is as a distribution with a few types (as modelled in Example 1), where we have twelve mass clusters  $H = \cup_{k=0}^{k=5} \{k/24, (19+k)/24\}$  corresponding to the twelve points on the which the distribution is supported with eleven gap intervals  $G = \left( \cup_{k=0}^{k=4} (k/24, (k+1)/24) \cup ((19+k)/24, (20+k)/24) \right) \cup (5/24, 19/24)$ . We can easily verify that the distribution  $p_1(x)$  satisfies the conditions in Definition 2 with  $\beta = 0, \varepsilon_0 = \min_{\{x:p_1(x)>0\}} p_1(x) = 1/42$  and  $\delta = 0$ . The second way to model this distribution is by having only two mass clusters  $H_1 = [0, 1/4]$  and  $H_2 = [3/4, 1]$  with one gap interval  $G = (1/4, 3/4)$ . By considering only two clusters, we have that  $\varepsilon_0 = 1/2$  since the total probability

mass in both the clustered is  $1/2$  each. It is easy to see that for any choice of  $\beta \in [0, \infty)$ , to satisfy condition (a) in Definition 2, we must choose  $\delta = \eta_1 > 0$ . While both ways are valid in terms of modelling the distribution, the theoretical guarantees implied by the two different characterizations of the same distribution lead to two different regret scalings. Under the first way where  $p_1$  is modelled as a  $(\beta = 0, \varepsilon_0 = 1/42, \delta = 0)$ -clustered, we get constant regret scaling, while under the second way where  $p_1$  is modelled as a  $(\beta = 0, \varepsilon_0 = 1/2, \delta = 1/24)$ -clustered, we get that the regret will scale as  $\tilde{O}(\sqrt{T})$ . Note that these regret scalings not only follow from the bounds in Theorem 5 but also due to the fact that CWG algorithm in Algorithm 2 will operate differently under the two different characterizations of the same distribution  $p_1$  since the gaps are defined differently under the two different characterizations.

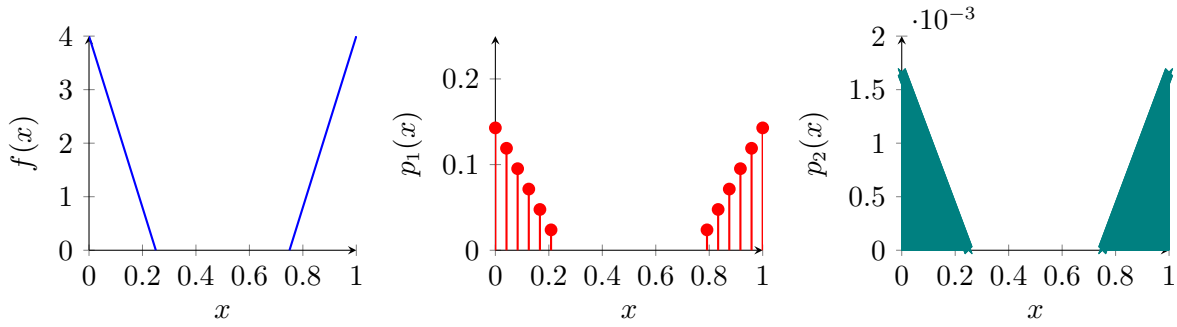


Figure B.4: (Left) PDF  $f_\beta$  of distribution  $F$  (Center) a few types (Right) many small types

Coming to the distribution  $p_2$ , again there are two ways that the distribution  $p_2$  can be modelled as a  $(\beta, \varepsilon_0, \delta)$ -clustered distribution. Since strictly speaking,  $p_2$  is an atomic distribution albeit with many types, we can model it similar to how we modelled an atomic distribution with a few types. Building on that, we would have that 1200 mass clusters  $H = \cup_{k=0}^{598} \{k/2400, (1801 + k)/2400\}$  with 1199 gap intervals  $G = \left( \cup_{k=0}^{598} (k/2400, (k + 1)/2400) \cup ((1801 + k)/2400, (1802 + k)/2400) \right) \cup (599/2400, 1801/2400)$ . We can easily verify that the distribution  $p_2(x)$  satisfies the conditions in Definition 2 with  $\beta = 0, \varepsilon_0 = \min_{\{x: p_2(x) > 0\}} p_2(x) = 1/360600$  and  $\delta = 0$ . The second way to model this distribution is having only two mass clusters  $H_1 = [0, 1/4]$  and  $H_2 = [3/4, 1]$  with one gap interval  $G = (1/4, 3/4)$ . By considering only two clusters, we have that  $\varepsilon_0 = 1/2$ . It is easy to verify that for  $\beta = 1$  and  $\delta = \eta_2$ , we satisfy the condition (a) in Definition 2. Note that under

the first way, we have that  $\varepsilon_0$  is very small and for most reasonable and practical values of the time horizon  $T$ , we may have that  $1/\varepsilon_0 \sim T$  and hence the theoretical guarantees implied by Theorem 5 may be vacuous. On the other hand, in the second characterization as  $(\beta = 1, \varepsilon_0 = 1/2, \delta = \eta_2)$ -clustered distribution, we have that  $\delta \sim 1/\sqrt{T}$  for reasonable values of  $T$  and implied regret scaling is  $\tilde{O}(T^{\frac{1}{4}})$  (sublinear regret). Note that the CWG algorithm (Algorithm 2) operates differently under the two different characterizations of the same distribution  $p_2$ .

## Appendix C: The Fault in Our Recommendations: On the Perils of Optimizing the Measurable

### C.1 Useful Technical Result

The following proposition characterizes the probability of engagement and the expected utility of the user given the user is recommended items  $\pi = \{i_1, i_2\}$ .

**Proposition 5** *Given a recommendation  $\pi = \{i_1, i_2\}$ , the expected engagement and utility is given as*

$$\begin{aligned} \mathbb{E} [\mathbb{1}\{c(\pi) \neq \emptyset\}] &= \mathbb{P}(c(\pi) \neq \emptyset) = \frac{e^{V_{\tau(i_1)}} + e^{V_{\tau(i_2)}}}{1 + e^{V_{\tau(i_1)}} + e^{V_{\tau(i_2)}}} \\ \mathbb{E} \left[ \max_{j \in \{i_1, i_2, \emptyset\}} u_j \right] &= \ln \left( 1 + e^{V_{\tau(i_1)}} + e^{V_{\tau(i_2)}} \right) \end{aligned}$$

For a proof of Proposition 5, refer to [120, Chapter 2].

### C.2 Proof of Theorem 7

*Proof of Theorem 7.* We fix  $\delta \in [0, 1)$  and the attraction parameter of the popular type  $V_P \in \mathbb{R}_+$ .

The expected engagement under APP is given as

$$\text{Eng}(\text{APP}) = \frac{2e^{V_P}}{1 + 2e^{V_P}} + \frac{\delta}{1 - \delta} \cdot \frac{2e^{V_P}}{1 + 2e^{V_P}} \quad (\text{C.1})$$

Let E-OPT denote the engagement optimal policy in the class of all online policies  $\Pi$ . Consider the following exploration-based policy denoted Diverse-then-Optimal (DO) – it recommends one popular and one niche type at the first time step and from the second step onwards, implements

E-OPT. The expected engagement under DO is given as

$$\begin{aligned} \text{Eng}(\text{DO}) &= (1-p) \cdot \frac{e^{V_P} + e^{-1}}{1 + e^{V_P} + e^{-1}} + p \cdot \frac{e^{V_P} + e^{(1-p)/p}}{1 + e^{V_P} + e^{(1-p)/p}} + \frac{\delta}{1-\delta} \cdot \text{Eng}(\text{E-OPT}) \\ &\leq (1-p) \cdot \frac{e^{V_P} + e^{-1}}{1 + e^{V_P} + e^{-1}} + p \cdot \frac{e^{V_P} + e^{(1-p)/p}}{1 + e^{V_P} + e^{(1-p)/p}} + \frac{\delta}{1-\delta} \left[ p \cdot \frac{2e^{(1-p)/p}}{1 + 2e^{(1-p)/p}} + (1-p) \cdot \frac{2e^{V_P}}{1 + 2e^{V_P}} \right] \end{aligned}$$

where the inequality follows from the fact that the engagement under the optimal online policy is bounded above by the engagement of an oracle who knows whether a user prefers niche to popular type and recommends the preferred type. Then we have that

$$\lim_{p \rightarrow 0} [\text{Eng}(\text{APP}) - \text{Eng}(\text{DO})] = \frac{2e^{V_P}}{1 + 2e^{V_P}} - \frac{e^{V_P} + e^{-1}}{1 + e^{V_P} + e^{-1}} > 0$$

Hence the limit  $p \rightarrow 0$ , we have that APP does better than the best exploration-driven policy DO and hence it is optimal. The expected utility under APP is  $\frac{1}{1-\delta} \cdot \ln(1 + 2e^{V_P})$  which follows as a corollary of Proposition 5. ■

### C.3 Proof of Theorem 8

*Proof of Theorem 8.* We begin by defining some quantities. Fix a discount factor  $\delta \in [0, 1)$  and the parameter  $p$  from (4.5). Define the following

$$\rho_1 \triangleq \frac{e^{(1-p)/p}}{1 + e^{V_P} + e^{(1-p)/p}}, \quad \rho_2 \triangleq \frac{e^{-1}}{1 + e^{V_P} + e^{-1}}, \quad c \triangleq \frac{\ln((1-\rho_2)/(1-\rho_1))}{\ln(\rho_1/\rho_2) + \ln((1-\rho_2)/(1-\rho_1))} \quad (\text{C.2})$$

For a  $\rho, x \in [0, 1]$ , define the following random variable

$$X_k(\rho, x) = \begin{cases} 1-x, & \text{with probability } \rho \\ -x, & \text{with probability } 1-\rho \end{cases} \quad (\text{C.3})$$

Let  $S_n(\rho, x) = \sum_{k=1}^n X_k(\rho, x)$  denote the  $n$ -th partial sum of the random walk described by  $X_k(\rho, x)$ .

Define  $N(\rho, x)$  to be the first time the partial sum goes below zero, i.e.,

$$N(\rho, x) = \inf\{n : S_n(\rho, x) < 0\} \quad (\text{C.4})$$

PEAR recommends one item of popular type and another item of niche type while the posterior belief (denoted as  $p_t$ ) on  $V_N = (1 - p)/p$  is greater than or equal to  $p$ . If the posterior  $p_t < p$ , then PEAR switches to showing both items of the popular type. The time at each the policy switches from showing one popular and one niche type to both popular type is given by the random variable  $N(\rho, c)$  defined in (C.4) for different values of  $\rho$  and this is characterized in the following lemma.

**Lemma 20** *Consider  $\rho_1, \rho_2$  and  $c$  defined in (C.2). If  $V_N = (1 - p)/p$ , then the first time PEAR recommends both the items of popular type is  $N(\rho_1, c)$ . Analogously, if  $V_N = -1$ , then the first time PEAR recommends both the items of the popular type is  $N(\rho_2, c)$ .*

We defer the proof of Lemma 20 to Appendix C.4.1. Lemma 20 implies that from  $t = 0$  to  $t = N(\rho, c) - 1$ , PEAR recommends one popular and one niche type item and for  $t \geq N(\rho, c)$ , PEAR recommends both items of the popular type. Hence the expected engagement under PEAR is given as

$$\begin{aligned} \text{Eng(PEAR)} &= p \cdot \mathbb{E} \left[ \beta_1 \sum_{t=0}^{N(\rho_1, c)-1} \delta^t + \lambda \sum_{t=N(\rho_1, c)}^{\infty} \delta^t \right] + (1 - p) \cdot \mathbb{E} \left[ \beta_2 \sum_{t=0}^{N(\rho_2, c)-1} \delta^t + \lambda \sum_{t=N(\rho_2, c)}^{\infty} \delta^t \right] \\ &= \frac{p}{1 - \delta} \cdot (\beta_1 g(\delta, \rho_1, c) + \lambda(1 - g(\delta, \rho_1, c))) + \frac{1 - p}{1 - \delta} \cdot (\beta_2 g(\delta, \rho_2, c) + \lambda(1 - g(\delta, \rho_2, c))), \end{aligned}$$

where  $\beta_1 = \frac{e^{V_P + e^{(1-p)/p}}}{1 + e^{V_P + e^{(1-p)/p}}}$ ,  $\beta_2 = \frac{e^{V_P + e^{-1}}}{1 + e^{V_P + e^{-1}}}$ ,  $\lambda = \frac{2e^{V_P}}{1 + 2e^{V_P}}$  and  $g(\delta, \rho, x)$  is defined as follows

$$g(\delta, \rho, x) \triangleq \mathbb{E} [1 - \delta^{N(\rho, x)}]. \quad (\text{C.5})$$

Similarly, the expected utility under PEAR is given as

$$\begin{aligned} \text{Util(PEAR)} &= p \cdot \mathbb{E} \left[ \Psi_1 \sum_{t=0}^{N(\rho_1, c)-1} \delta^t + \Lambda \sum_{t=N(\rho_1, c)}^{\infty} \delta^t \right] + (1-p) \cdot \mathbb{E} \left[ \Psi_2 \sum_{t=0}^{N(\rho_2, c)-1} \delta^t + \Lambda \sum_{t=N(\rho_2, c)}^{\infty} \delta^t \right] \\ &= \frac{p}{1-\delta} \cdot (\Psi_1 g(\delta, \rho_1, c) + \Lambda(1 - g(\delta, \rho_1, c))) + \frac{1-p}{1-\delta} \cdot (\Psi_2 g(\delta, \rho_2, c) + \Lambda(1 - g(\delta, \rho_2, c))), \end{aligned}$$

where  $\Psi_1 = \ln(1 + e^{V_P} + e^{(1-p)/p})$ ,  $\Psi_2 = \ln(1 + e^{V_P} + e^{-1})$  and  $\Lambda = \ln(1 + 2e^{V_P})$ . Finally, we are interested in the limit  $p \rightarrow 0$ . Note that  $g(\delta, \rho_1, c)$  and  $g(\delta, \rho_2, c)$  are also function of  $p$ . The following lemma characterize the limiting value of  $g(\delta, \rho_1, c)$  and  $g(\delta, \rho_2, c)$  when  $p \rightarrow 0$ . We defer the proof of Lemma 21 to Appendix C.4.2.

**Lemma 21** Fix  $\delta \in [0, 1)$  and consider  $\rho_1, \rho_2$  and  $c$  defined in (C.2). Then  $\lim_{p \rightarrow 0} g(\delta, \rho_1, c) = 1$  and  $\lim_{p \rightarrow 0} g(\delta, \rho_2, c) = \frac{1-\delta}{1-\delta\rho_2}$ .

In the limit  $p \rightarrow 0$ , the expected engagement and utility is

$$\begin{aligned} \lim_{p \rightarrow 0} \text{Eng(PEAR)} &= \frac{1}{1-\delta} \left( \beta_2 \cdot \frac{1-\delta}{1-\delta\rho_2} + \lambda \cdot \frac{\delta(1-\rho_2)}{1-\delta\rho_2} \right) \\ \lim_{p \rightarrow 0} \text{Util(PEAR)} &= \frac{1}{1-\delta} \left( 1 + \Psi_2 \cdot \frac{1-\delta}{1-\delta\rho_2} + \Lambda \cdot \frac{\delta(1-\rho_2)}{1-\delta\rho_2} \right) \end{aligned}$$

This concludes the proof. ■

## C.4 Proof of Helper Lemmas

In this section, we provide the proof of some helper lemmas used in the proof of Theorem 8.

### C.4.1 Proof of Lemma 20

*Proof of Lemma 20.* Note that PEAR switches to showing both the items of the popular type whenever  $p_t < p$  where  $p_t$  is the posterior belief of  $V_N = p/(1-p)$ . Note that  $p_t$  depends on the (random) number of successes  $S$  and failures  $F$  observed till time  $t$ , where if we count the user choosing the niche item as success and the user not choosing the niche item (i.e. either choosing

the popular item or the outside option) as failure. We have that the following are equivalent

$$\begin{aligned}
p_t < p &\stackrel{(a)}{\iff} \frac{1}{p} < \frac{1-p}{p} \frac{\rho_2^S(1-\rho_2)^F}{\rho_1^S(1-\rho_1)^F} + 1 \\
&\stackrel{(b)}{\iff} \rho_1^S(1-\rho_1)^F < \rho_2^S(1-\rho_2)^F \\
&\stackrel{(c)}{\iff} S \left( \frac{\ln\left(\frac{\rho_1}{\rho_2}\right)}{\ln\left(\frac{\rho_1}{\rho_2}\right) + \ln\left(\frac{1-\rho_2}{1-\rho_1}\right)} \right) + F \left( -\frac{\ln\left(\frac{1-\rho_2}{1-\rho_1}\right)}{\ln\left(\frac{\rho_1}{\rho_2}\right) + \ln\left(\frac{1-\rho_2}{1-\rho_1}\right)} \right) < 0 \\
&\stackrel{(d)}{\iff} S(1-c) + F(-c) < 0,
\end{aligned}$$

where (a) follows from the definition of  $p_t$ , (b) and (c) follows from algebraic manipulations and (d) follows from the definition of  $c$  in (C.2). Note that the equation  $S(1-c) + F(-c)$  corresponds to the  $S + F$ -th partial sum of a random walk with steps of size  $1-c$  or  $-c$ , which is the same random walk as described in (C.3). Therefore  $N(\rho_1, c)$  and  $N(\rho_2, c)$  correspond to the stopping time when  $V_N = p/(1-p)$  and  $V_N = -1$ . ■

#### C.4.2 Proof of Lemma 21

*Proof of Lemma 21.* Define  $T(\rho, c) = \inf\{n : X_n(\rho, c) = -c\}$  for  $\rho \in \{\rho_1, \rho_2\}$  and  $M_0 = \frac{\ln((1-\rho_2)/(1-\rho_1))}{\ln(\rho_1/\rho_2)} = \Theta(1/p)$ . Recall the random variable  $N(\rho, x)$  is defined in (C.4). Now for all  $k \leq M_0$ , we have that the following events are equivalent,

$$\{T(\rho, c) = k\} \equiv \{N(\rho, c) = k\} \text{ for } \rho \in \{\rho_1, \rho_2\} \quad (\text{C.6})$$

We first show that  $\lim_{p \rightarrow 0} g(\delta, \rho_1, c) = 1$ . Since  $\delta \in (0, 1]$ , we trivially have that  $g(\delta, \rho_1, c) \leq 1$ . Next we will lower bound  $g(\delta, \rho_1, c)$ .

$$\begin{aligned}
g(\delta, \rho_1, c) &\stackrel{(a)}{=} 1 - \sum_{k=1}^{\infty} \delta^k \mathbb{P}(N(\rho_1, c) = k) \\
&\stackrel{(b)}{=} 1 - \sum_{k=1}^{M_0} \delta^k \mathbb{P}(T(\rho_1, c) = k) - \sum_{k=M_0+1}^{\infty} \delta^k \mathbb{P}(N(\rho_1, c) = k) \\
&\stackrel{(c)}{\geq} 1 - \frac{1 - \rho_1}{\rho_1} \sum_{k=1}^{M_0} \delta^k \rho_1^k - \delta^{M_0+1} \\
&\stackrel{(d)}{=} 1 - \frac{\delta(1 - \rho_1)(1 - (\delta\rho_1)^{M_0})}{1 - \delta\rho_1} - \delta^{M_0+1} \\
&\stackrel{(e)}{\geq} 1 - e^{-1/p} \frac{e\delta(1 + e^{V_P})}{1 - \delta} - \delta^{\Theta(1/p)}
\end{aligned}$$

where (a) follows from the definition of  $g(\delta, \rho, x)$  in (C.5), (b) follows from (C.6) for  $\rho = \rho_1$ , (c) follows from the fact that  $\mathbb{P}(T(\rho_1, c) = k) = (1 - \rho_1)\rho_1^{k-1}$  and the fact that  $\delta^{M_0} \geq \delta^k$  for all  $k \geq M_0$ , (d) follows from sum of geometric series, (e) follows from the fact that  $1 - \rho_1 \leq (1 + e^{V_P})e^{-1/p+1}$ ,  $1 - (\delta\rho_1)^{M_0} \leq 1$  and  $1 - \delta\rho_1 \geq 1 - \delta$ . Taking the limit of  $p \rightarrow 0$  for a fixed  $\delta \in [0, 1)$  and  $V_P \in \mathbb{R}_+$ , we have that

$$\lim_{p \rightarrow 0} g(\delta, \rho_1, c) \geq \lim_{p \rightarrow 0} \left( 1 - e^{-1/p} \frac{e\delta(1 + e^{V_P})}{1 - \delta} - \delta^{\Theta(1/p)} \right) = 1$$

Next we want that  $\lim_{p \rightarrow 0} g(\delta, \rho_2, c) = \frac{1 - \delta}{1 - \delta\rho_2}$ . In particular, we will show that

$$\frac{1 - \delta}{1 - \delta\rho_2} + \frac{e(1 + e^{V_P})(\delta\rho_2)^{M_0+1}}{1 - \delta\rho_2} - \delta^{M_0+1} \leq g(\delta, \rho_2, c) \leq \frac{1 - \delta}{1 - \delta\rho_2} + \frac{e(1 + e^{V_P})(\delta\rho_2)^{M_0+1}}{1 - \delta\rho_2}$$

We will begin with the upper bound.

$$\begin{aligned}
g(\delta, \rho_2, c) &\stackrel{(a)}{=} 1 - \sum_{k=1}^{\infty} \delta^k \mathbb{P}(N(\rho_2, c) = k), \\
&\stackrel{(b)}{\leq} 1 - \sum_{k=1}^{M_0} \delta^k \mathbb{P}(T(\rho_2, c) = k), \\
&\stackrel{(c)}{=} 1 - \frac{1 - \rho_2}{\rho_2} \sum_{k=1}^{M_0} \delta^k \rho_2^k, \\
&\stackrel{(d)}{=} 1 - e(1 + e^{V_P}) \frac{\delta \rho_2 (1 - (\delta \rho_2)^{M_0})}{1 - \delta \rho_2}, \\
&\stackrel{(e)}{=} \frac{1 - \delta \rho_2 - e(1 + e^{V_P}) \delta \rho_2 + e(1 + e^{V_P}) (\delta \rho_2)^{M_0+1}}{1 - \delta \rho_2}, \\
&\stackrel{(f)}{=} \frac{1 - \delta}{1 - \delta \rho_2} + \frac{e(1 + e^{V_P}) (\delta \rho_2)^{M_0+1}}{1 - \delta \rho_2},
\end{aligned}$$

where (a) follows from the definition of  $g(\delta, \rho, x)$  in (C.5), (b) follows from (C.6) for  $\rho = \rho_2$ , (c) follows from the fact that  $\mathbb{P}(T(\rho_2, c) = k) = (1 - \rho_2) \rho_2^{k-1}$ , (d) follows from sum of the geometric series, (e) follows trivially, (f) follows from the fact that  $e(1 + e^{V_P}) \delta \rho_2 = (1 - \rho_2)$ . The lower bound on  $g(\delta, \rho_2, c)$  follows using a similar line of argument. Note that as  $p \rightarrow 0$ , since  $\delta \rho_2 < 1$  and  $\delta < 1$ , we have that  $(\delta \rho_2)^{M_0+1} \rightarrow 0$  and  $\delta^{M_0+1} \rightarrow 0$  and together we have that  $\lim_{p \rightarrow 0} g(\delta, \rho_2, c) = \frac{1 - \delta}{1 - \delta \rho_2}$ . ■

## Appendix D: Impact of Rankings and Personalized Recommendations in Marketplaces

### D.1 Proof of intermediate results

#### D.1.1 Proof of Proposition 4

*Proof of Proposition 4.* Fix  $\epsilon > 0$ . Since  $X$  has a pareto tail with parameters  $c > 0$  and  $\alpha > 1$ , there exists a constant  $x_0 > 0$  such that for all  $x \geq x_0$ , we have that

$$(1 - \epsilon)(c/x)^\alpha \leq \mathbb{P}(X > x) \leq (1 + \epsilon)(c/x)^\alpha. \quad (\text{D.1})$$

Next we want to bound the tail distribution for  $X_{(n:n)}$  which is maximum of  $n$  i.i.d copies of  $X$ . We have that for all  $x \geq x_0$ , we have that

$$\mathbb{P}(X_{(n:n)} > x) = 1 - \mathbb{P}(X_{(n:n)} \leq x) = 1 - (\mathbb{P}(X \leq x))^n = 1 - (1 - \mathbb{P}(X > x))^n. \quad (\text{D.2})$$

Using (D.1) and (D.2), we have that for  $x \geq x_0$ ,

$$1 - (1 - (1 - \epsilon)(c/x)^\alpha)^n \leq \mathbb{P}(X_{(n:n)} > x) \leq 1 - (1 - (1 + \epsilon)(c/x)^\alpha)^n. \quad (\text{D.3})$$

Since  $X \geq 0$ , we have that  $X_{(n:n)} \geq 0$  and therefore, we will make use of the tail sum formula for the expectation to provide upper and lower bounds on  $\mathbb{E}[X_{(n:n)}]$ . Define  $\bar{c} \triangleq (1 + \epsilon)^{1/\alpha}c$  and  $\underline{c} \triangleq (1 - \epsilon)^{1/\alpha}c$ . We will begin by providing the upper bound on  $\mathbb{E}[X_{(n:n)}]$ .

$$\mathbb{E}[X_{(n:n)}] \stackrel{(a)}{=} \int_0^\infty \mathbb{P}(X_{(n:n)} > x) dx \stackrel{(b)}{\leq} \max\{x_0, \bar{c}\} + \int_{\bar{c}}^\infty 1 - (1 - (\bar{c}/x)^\alpha)^n dx, \quad (\text{D.4})$$

where (a) follows from the tail sum formula for the expectation [164] and (b) follows from the fact that in the interval  $[0, \max\{x_0, \bar{c}\}]$ , we have that  $\mathbb{P}(X_{(n:n)} > x) \leq 1$  and  $\int_{\max\{x_0, \bar{c}\}}^{\infty} 1 - (1 - (c/x)^\alpha)^n dx \leq \int_{\bar{c}}^{\infty} 1 - (1 - (\bar{c}/x)^\alpha)^n dx$ .

Next we will compute the integral  $\int_{\bar{c}}^{\infty} 1 - (1 - (\bar{c}/x)^\alpha)^n dx$ . Let  $U(x) = x$  and  $V(x) = 1 - (1 - (\bar{c}/x)^\alpha)^n$ . Therefore  $dU(x) = dx$  and  $dV(x) = -n\alpha\bar{c}^\alpha(1 - (\bar{c}/x)^\alpha)^{n-1}x^{-\alpha-1} dx$ .

$$\begin{aligned}
\int_{\bar{c}}^{\infty} 1 - (1 - (\bar{c}/x)^\alpha)^n dx &\stackrel{(a)}{=} \int_{\bar{c}}^{\infty} V(x)dU(x), \\
&\stackrel{(b)}{=} [U(x)V(x)]_{\bar{c}}^{\infty} - \int_{\bar{c}}^{\infty} U(x)dV(x), \\
&\stackrel{(c)}{=} -\bar{c} - \int_{\bar{c}}^{\infty} x \cdot (-n\alpha\bar{c}^\alpha(1 - (\bar{c}/x)^\alpha)^{n-1}x^{-\alpha-1})dx, \\
&\stackrel{(d)}{=} -\bar{c} + n\alpha \int_{\bar{c}}^{\infty} (1 - (\bar{c}/x)^\alpha)^{n-1}(\bar{c}/x)^\alpha dx, \\
&\stackrel{(e)}{=} -\bar{c} + n\bar{c} \int_0^1 (1-u)^{n-1}u^{-\frac{1}{\alpha}+1-1} du, \\
&\stackrel{(f)}{=} -\bar{c} + n\bar{c} \frac{\Gamma(1-1/\alpha)\Gamma(n)}{\Gamma(n+1-1/\alpha)}, \tag{D.5}
\end{aligned}$$

where (a) follows from the definition of  $U(x)$  and  $V(x)$ , (b) follows from integration by parts, (c) follows from  $dV(x)$ , (d) follows from rearrangement, (e) follows from change of variable where  $(\bar{c}/x)^\alpha = u$  and simplification, (f) from the the definition of Beta function  $B(t_1, t_2) = \int_0^1 u^{t_1-1}(1-u)^{t_2-1} du = \Gamma(t_1)\Gamma(t_2)/\Gamma(t_1+t_2)$  where  $t_1 = 1 - 1/\alpha$  and  $t_2 = n$ . Combining (D.4) and (D.5), we have that

$$\mathbb{E}[X_{(n:n)}] \leq (x_0 - \bar{c})_+ + n\bar{c} \frac{\Gamma(1-1/\alpha)\Gamma(n)}{\Gamma(n+1-1/\alpha)}. \tag{D.6}$$

Using Stirlings' approximation, we have that

$$\lim_{n \rightarrow \infty} \frac{n\Gamma(n)}{\Gamma(n+1-1/\alpha)n^{1/\alpha}} = 1 \tag{D.7}$$

Combining (D.6) and (D.7), we have that

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\Gamma(1 - 1/\alpha)n^{1/\alpha}} \leq \limsup_{n \rightarrow \infty} \left\{ \frac{1}{\Gamma(1 - 1/\alpha)n^{1/\alpha}} \cdot n\bar{c} \frac{\Gamma(1 - 1/\alpha)\Gamma(n)}{\Gamma(n + 1 - 1/\alpha)} \right\} = \bar{c}.$$

Using similar arguments as provided for the upper bound, we can easily show the following lower bound,

$$\underline{c} = \liminf_{n \rightarrow \infty} \left\{ \frac{1}{\Gamma(1 - 1/\alpha)n^{1/\alpha}} \cdot n\underline{c} \frac{\Gamma(1 - 1/\alpha)\Gamma(n)}{\Gamma(n + 1 - 1/\alpha)} \right\} \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\Gamma(1 - 1/\alpha)n^{1/\alpha}}.$$

Combining these two results along with the definition of  $\underline{c} = (1 - \epsilon)^{1/\alpha}c$  and  $\bar{c} = (1 + \epsilon)^{1/\alpha}c$ , we have that

$$(1 - \epsilon)^{1/\alpha} \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{c\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} \leq \limsup_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{c\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} \leq (1 + \epsilon)^{1/\alpha}$$

Note that since the above set of inequalities hold for every  $\epsilon > 0$ , we have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{c\Gamma(1 - 1/\alpha) \cdot n^{1/\alpha}} = 1.$$

This completes the proof. ■

### D.1.2 Proof of Lemma 3

*Proof of Lemma 3.* Since  $X \geq 0$  and  $Y \geq 0$ , it trivially follows that  $Z \geq 0$  and from linearity of expectations we have that  $\mathbb{E}[Z] = (1 - \rho)\mu_X + \rho\mu_Y < \infty$ .

Since  $X$  has Pareto tail with parameters  $c_X > 0$  and  $\alpha_X > 1$ , we have that  $(1 - \rho)X$  has a pareto tail with parameters  $(1 - \rho)c_X > 0$  and  $\alpha_x > 1$ . This is because

$$1 = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{(c_X/x)^\alpha} = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x/(1 - \rho))}{(c_X/(x/(1 - \rho)))^\alpha} = \lim_{x \rightarrow \infty} \frac{\mathbb{P}((1 - \rho)X > x)}{((1 - \rho)c_X/x)^\alpha}$$

Similarly we have that  $\rho Y$  has a pareto tail with parameters  $\rho c_Y > 0$  and  $\alpha_Y > 1$ . Let us denote

$\tilde{X} = (1 - \rho)X$  and  $\tilde{Y} = \rho Y$ , then  $Z = \tilde{X} + \tilde{Y}$  and we want to characterize the tail behavior of  $Z$ .

We will begin by providing a lower bound on the tail of  $Z$ . We have that

$$\begin{aligned}
\mathbb{P}(Z > t) &\stackrel{(a)}{=} \mathbb{P}(\tilde{X} + \tilde{Y} > t), \\
&\stackrel{(b)}{\geq} \mathbb{P}(\max\{\tilde{X}, \tilde{Y}\} > t), \\
&\stackrel{(c)}{=} 1 - (1 - \mathbb{P}(\tilde{X} > t))(1 - \mathbb{P}(\tilde{Y} > t)), \\
&\stackrel{(d)}{=} \mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t) - \mathbb{P}(\tilde{X} > t)\mathbb{P}(\tilde{Y} > t), \tag{D.8}
\end{aligned}$$

where (a) follows from the definition of  $Z$ , (b) follows from the fact that  $\{\max\{\tilde{X}, \tilde{Y}\} > t\} \implies \{Z > t\}$ , (c) follows from the fact that  $\mathbb{P}(\max\{\tilde{X}, \tilde{Y}\} > t) = 1 - \mathbb{P}(\tilde{X} < t)\mathbb{P}(\tilde{Y} < t)$  since  $\tilde{X}$  and  $\tilde{Y}$  are independent, (d) follows trivially. Next we will provide an upper bound on the tail of  $Z$ . Fix a  $\delta \in (0, 1/2)$ . We have that

$$\mathbb{P}(Z > t) = \mathbb{P}(\tilde{X} + \tilde{Y} > t) \stackrel{(a)}{\leq} \mathbb{P}(\tilde{X} > (1 - \delta)t) + \mathbb{P}(\tilde{Y} > (1 - \delta)t) + \mathbb{P}(\tilde{X} > \delta t)\mathbb{P}(\tilde{Y} > \delta t), \tag{D.9}$$

where (a) follows from the fact that the event  $\{\tilde{X} + \tilde{Y} > t\}$  implies the event  $\{\tilde{X} > (1 - \delta)t\} \cup \{\tilde{Y} > (1 - \delta)t\} \cup \{\tilde{X} > \delta t, \tilde{Y} > \delta t\}$ .

Next we will consider following cases:

(a)  $\alpha_X < \alpha_Y$ . Using (D.8), we have that  $\liminf_{t \rightarrow \infty} \mathbb{P}(Z > t)/\mathbb{P}(\tilde{X} > t) \geq 1$  and using (D.9), we have that  $\limsup_{t \rightarrow \infty} \mathbb{P}(Z > t)/\mathbb{P}(\tilde{X} > t) \leq \limsup_{t \rightarrow \infty} \mathbb{P}(\tilde{X} > (1 - \delta)t)/\mathbb{P}(\tilde{X} > t) = (1 - \delta)^{-\alpha_X}$ . Since the upper bound holds for all  $\delta \in (0, 1/2)$ , we have that  $\lim_{t \rightarrow \infty} \mathbb{P}(Z > t)/\mathbb{P}(\tilde{X} > t) = 1$ . Therefore we have that  $Z$  has a pareto tail with parameters  $c_Z = (1 - \rho)c_X$  and  $\alpha_Z = \alpha_X$ .

(b)  $\alpha_X > \alpha_Y$ . This is completely analogous to the case above.

(c)  $\alpha_X = \alpha_Y = \alpha$ . Note that

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t)}{(c_Z/t)^\alpha} = 1, \quad \text{where } c_Z = (((1 - \rho)c_X)^\alpha + (\rho c_Y)^\alpha)^{1/\alpha}$$

Using (D.8), we have that  $\liminf_{t \rightarrow \infty} \mathbb{P}(Z > t)/(\mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t)) \geq 1$  and using (D.9), we have that  $\limsup_{t \rightarrow \infty} \mathbb{P}(Z > t)/(\mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t)) \leq \limsup_{t \rightarrow \infty} \mathbb{P}(\tilde{X} > (1 - \delta)t)/(\mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t)) = (1 - \delta)^{-\alpha}$ . Since the upper bound holds for all  $\delta \in (0, 1/2)$ , we have that  $\lim_{t \rightarrow \infty} \mathbb{P}(Z > t)/(\mathbb{P}(\tilde{X} > t) + \mathbb{P}(\tilde{Y} > t)) = 1$ . Therefore we have that  $Z$  has a pareto tail with parameters  $c_Z = (((1 - \rho)c_X)^\alpha + (\rho c_Y)^\alpha)^{1/\alpha}$  and  $\alpha_Z = \alpha$ .

This completes the proof. ■

### D.1.3 Proof of Lemma 4

*Proof of Lemma 4.* Note that  $\Delta_{q \rightarrow u}^{\text{cap}}(n) = \text{AW}_u^{\text{cap}}(n) - \text{AW}_q^{\text{cap}}(n)$ . We will begin by characterizing the social welfare in the Only Quality Information regime. In the Only Quality Information regime, the agents base their decisions solely on the common term ( $q_y$ ). Therefore, we have that the agent  $k$  will choose the item with common term value  $q_{(k:n)}$  (recall that  $X_{(k:n)}$  denotes the  $k$ -th smallest value of  $n$  i.i.d copies of  $X$ ). Recall that  $\sigma_q(k)$  denotes the index of the item chosen by agent  $k$  in the Only Quality Information regime. This means that  $q_{\sigma_q(k)} = q_{(k:n)}$ . Therefore we have that,

$$\begin{aligned} \text{AW}_q^{\text{cap}}(n) &\stackrel{(a)}{=} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n (1 - \rho) q_{\sigma_q(k)} + \rho \varphi_{k\sigma_q(k)} \right] \\ &\stackrel{(b)}{=} (1 - \rho) \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n q_k \right] + \rho \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\varphi_{k\sigma_q(k)}], \\ &\stackrel{(c)}{=} (1 - \rho) \mu_q + \rho \mu_\varphi, \end{aligned} \tag{D.10}$$

where (a) follows from definition of  $u_{k\sigma_q(k)}$ , (b) follows from the fact that  $\sum_{k=1}^n q_{\sigma_q(k)} = \sum_{k=1}^n q_{(k:n)} = \sum_{k=1}^n q_k$ , (c) follows from the fact that  $\mathbb{E}[q_k] = \mu_q$  and  $\mathbb{E}[\varphi_{k\sigma_q(k)}] = \mu_\varphi$  since the index  $\sigma_q(k)$  is

random and hence we have that  $\varphi_{k\sigma_0(k)}$  is a random sample drawn from the distribution  $P_\varphi$ .

Next we will provide an upper and lower bound on the social welfare in the Full Information regime. In the model description in Section 5.2, every agent in  $\mathcal{X}$  observes the common terms ( $q_y$ ) and the idiosyncratic terms ( $\varphi_{xy}$ ) for all items. However from an equivalent description is to have the agent  $k$  observe the idiosyncratic terms ( $\varphi_{xy}$ ) only for the *remaining* items. Since  $\varphi_{ky}$  are drawn i.i.d across agents, it is equivalent to assume that the idiosyncratic term  $\varphi_{ky}$  is drawn i.i.d for  $n - k$  items when it is agent  $k$ 's turn make the choice. This equivalence follows from the so-called "Principle of Deferred Decisions" [159].

We will first provide an upper bound on the social welfare  $\text{AW}_u^{\text{cap}}(n)$ . Recall that the  $\sigma_u(k)$  denotes the index of the item chosen by agent  $k$ . We have that

$$\begin{aligned}
\text{AW}_u^{\text{cap}}(n) &\stackrel{(a)}{=} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n (1 - \rho) q_{\sigma_u(k)} + \rho \varphi_{k\sigma_u(k)} \right], \\
&\stackrel{(b)}{=} (1 - \rho) \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n q_k \right] + \rho \frac{1}{n} \sum_{k=1}^n \mathbb{E} [\varphi_{k\sigma_u(k)}], \\
&\stackrel{(c)}{\leq} (1 - \rho) \mu_q + \rho n^{-1} \sum_{k=1}^n \mathbb{E} [\varphi_{k,(n-k:n-k)}], \\
&\stackrel{(d)}{=} (1 - \rho) \mu_q + \rho \Phi_n,
\end{aligned} \tag{D.11}$$

where (a) follows from definition of  $u_{k\sigma_u(k)}$ , (b) follows from the fact that  $\sum_{k=1}^n q_{\sigma_u(k)} = \sum_{k=1}^n q_k$ , (c) follows from the fact that  $\mathbb{E}[q_k] = \mu_q$  and  $\varphi_{k\sigma_u(k)} \leq \varphi_{k,(n-k:n-k)}$  where  $\varphi_{k,(n-k:n-k)}$  denotes the maximum of  $n - k$  i.i.d draws from  $P_\varphi$  for agent  $k$  and (d) follows from the definition of  $\Phi_n$ .

We will now present a lower bound on the social welfare  $\text{AW}_u^{\text{cap}}(n)$ . We have that

$$\text{AW}_u^{\text{cap}}(n) = \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n u_{k\sigma_u(k)} \right] \stackrel{(a)}{\geq} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^n \rho \varphi_{k,(n-k:n-k)} \right] \stackrel{(b)}{=} \rho \Phi_n, \tag{D.12}$$

where (a) follows from the fact that  $u_{k\sigma_u(k)} = \max_{y \in \mathcal{Y}_k^{\text{rem}}} (1 - \rho) q_y + \rho \varphi_{ky} \geq \rho \varphi_{k,(n-k:n-k)}$  since  $q_k \geq 0$  for all  $k$ , (b) follows from the definition of  $\Phi_n$ . Combining (D.10), (D.11) and (D.12) provides the required result. ■

## D.2 Proof of Theorems for utility distributions with Exponential tail

The result in Theorems 10 and 12 can be viewed as following from Theorems 9 and 11 respectively. See the informal discussion in D.2.1. For completeness, we provide a proof in Appendix D.2.3 and D.2.4 from first principles.

### D.2.1 Connection between Pareto and Exponential tail

It is well known that the exponential distribution is a special case of the generalized Pareto distribution [101]. In this section, we briefly discuss how the results in Theorem 9 and 11 can be used to derive the result in Theorem 10 and 12 under some appropriate joint scaling of the parameters and the market size  $n$ . First, it is useful to re-state the Pareto tail definition as  $\lim_{x \rightarrow \infty} \mathbb{P}(X > x)/(1+x/c)^{-\alpha} = 1$ , this is because  $\lim_{x \rightarrow \infty} (1+x/c)^{-\alpha}/(c/x)^\alpha = 1$ . Now, define  $\alpha = \ln n$ ,  $c = \ln n/\lambda$  and consider the double limit  $x \rightarrow \infty$  and  $n \rightarrow \infty$ . Now we have that

$$1 = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{\exp(-\lambda x)} = \lim_{x \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{P}(X > x)}{(1 + \lambda x / \ln n)^{-\ln n}} = \lim_{x \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{P}(X > x)}{(1 + x/c)^{-\alpha}} = \lim_{n \rightarrow \infty} \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{(1 + x/c)^{-\alpha}},$$

where the interchange of limit follows from the Moore-Osgood theorem [165].

As an illustration, we briefly explain how the result in Theorem (11.b) implies Theorem (12.b). Note that the denominator in Theorem (11.b) is actually  $C_\varphi \Gamma(n+1)/\Gamma(n+1-1/\alpha_\varphi)$  which simplifies to  $C_\varphi n^{1/\alpha_\varphi}$  using the Stirlings approximation. Note that constant  $C_\varphi = c_\varphi (\alpha_\varphi / (\alpha_\varphi + 1)) \Gamma(1 - 1/\alpha_\varphi)$ . Now plugging in  $\alpha_\varphi = \ln n$  and  $c_\varphi = \ln n/\lambda$  gives the result in Theorem (12.b) since  $\Gamma(1 - 1/\ln n) \xrightarrow{n \rightarrow \infty} \Gamma(1) = 1$ ,  $\ln n / (\ln n + 1) \xrightarrow{n \rightarrow \infty} 1$  and  $\Gamma(n+1)/\Gamma(n+1-1/\ln n) \xrightarrow{n \rightarrow \infty} 1$ . Similar idea can be use to derive Theorem 10 from Theorem 9.

### D.2.2 Useful Intermediate Results

The proof of Theorems 10 and 12 will make use of the following proposition which we state and prove below.

**Proposition 6** Let  $X$  be a random variable with distribution  $P$ . Assume that  $X \geq 0$  and  $\mathbb{E}[X] < \infty$ . Assume that  $X$  has an exponential tail with parameters  $c > 0$  and  $\lambda > 0$ . Let  $X_1, X_2, \dots, X_n$  be i.i.d copies of  $X$  and define  $X_{(n:n)} \triangleq \max_{1 \leq k \leq n} X_k$ . We have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\ln n / \lambda} = 1.$$

*Proof of Proposition 6.* Fix  $\epsilon > 0$ . Since  $X$  has an exponential tail with parameters  $c > 0$  and  $\lambda > 0$ , there exists a constant  $x_0 \geq 0$  such that for all  $x \geq x_0$ , we have that

$$(1 - \epsilon)c \exp(-\lambda x) \leq \mathbb{P}(X > x) \leq (1 + \epsilon)c \exp(-\lambda x) \quad (\text{D.13})$$

Next we want to bound the tail distribution for  $X_{(n:n)}$  which is maximum of  $n$  i.i.d copies of  $X$ . Define  $\bar{c} = (1 + \epsilon)c$  and  $\underline{c} = (1 - \epsilon)c$ . Using (D.2) and (D.13), we have that for  $x \geq x_0$ ,

$$1 - (1 - \underline{c} \exp(-\lambda x))^n \leq \mathbb{P}(X_{(n:n)} > x) \leq 1 - (1 - \bar{c} \exp(-\lambda x))^n$$

Since  $X \geq 0$ , we have that  $M \geq 0$  and therefore, we will make use of the tail sum formula for the expectation to provide upper and lower bounds on  $\mathbb{E}[X_{(n:n)}]$ . Define  $s \triangleq \ln \bar{c} / \lambda$ . We will begin by providing the upper bound on  $\mathbb{E}[X_{(n:n)}]$ .

$$\mathbb{E}[X_{(n:n)}] \stackrel{(a)}{=} \int_0^\infty \mathbb{P}(X_{(n:n)} > x) dx \stackrel{(b)}{\leq} \max\{x_0, s\} + \int_s^\infty 1 - (1 - \bar{c} \exp(-\lambda x))^n dx, \quad (\text{D.14})$$

where (a) follows from the tail sum formula for expectation [164], (b) follows from the fact that in the interval  $[0, \max\{x_0, s\}]$ , we have that  $\mathbb{P}(X_{(n:n)} > x) \leq 1$  and  $\int_{\max\{x_0, s\}}^\infty 1 - (1 - \bar{c} \exp(-\lambda x))^n dx \leq \int_s^\infty 1 - (1 - \bar{c} \exp(-\lambda x))^n dx$ .

We want to compute the integral  $\int_s^\infty 1 - (1 - \bar{c} \exp(-\lambda x))^n dx$ . Therefore, we have that,

$$\begin{aligned}
\int_s^\infty 1 - (1 - \bar{c} \exp(-\lambda x))^n dx &\stackrel{(a)}{=} \frac{1}{\lambda} \int_0^1 \frac{1 - (1 - u)^n}{u} du, \\
&\stackrel{(b)}{=} \frac{1}{\lambda} \int_0^1 \sum_{k=0}^{n-1} (1 - u)^k du, \\
&\stackrel{(c)}{=} \frac{1}{\lambda} \sum_{k=0}^{n-1} \int_0^1 (1 - u)^k du, \\
&\stackrel{(d)}{=} \frac{1}{\lambda} \sum_{k=0}^{n-1} \frac{1}{k+1}, \\
&\stackrel{(e)}{=} H_n / \lambda,
\end{aligned} \tag{D.15}$$

where (a) follows from the change of variable argument where  $u = \bar{c} \exp(-\lambda x)$  and some simplification, (b) follows from the fact that  $\frac{1-(1-u)^n}{u} = \sum_{k=0}^{n-1} (1-u)^k$ , (c) follows from the interchange between integral and summation, (d) follows from the fact that  $\int_0^1 (1-u)^k du = \frac{1}{k+1}$ , (e) follows from definition of harmonic number  $H_n = \sum_{k=1}^n \frac{1}{k} = \sum_{k=0}^{n-1} \frac{1}{k+1}$ .

It is easy to show that  $\lim_{n \rightarrow \infty} H_n / \ln n = 1$ . Therefore using (D.14) and (D.15), we have that

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\ln n / \lambda} \leq 1$$

Using similar arguments as provided for the upper bound, we can easily show that the following lower bound as well,

$$1 \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\ln n / \lambda}$$

Combining these two results, we have that  $\lim_{n \rightarrow \infty} \frac{\mathbb{E}[X_{(n:n)}]}{\ln n / \lambda} = 1$  and this completes the proof.  $\blacksquare$

**Proposition 7** Fix  $\rho \in (0, 1)$ . Let  $X$  be a random variable with non-negative support, finite mean  $\mu_X < \infty$  and has an exponential tail with parameters  $c_X > 0$  and  $\lambda_X > 0$ . Let  $Y$  be another random variable with non-negative support, finite mean  $\mu_Y < \infty$  and has an exponential tail with

parameters  $c_Y > 0$  and  $\lambda_Y > 0$ . Define  $Z = (1 - \rho)X + \rho Y$ . Let  $Z_{(n:n)} = \max\{Z_1, Z_2, \dots, Z_n\}$  where  $Z_1, Z_2, \dots, Z_n$  are i.i.d copies of  $Z$ . Then we have that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{\ln n} = \max \left\{ \frac{1 - \rho}{\lambda_X}, \frac{\rho}{\lambda_Y} \right\}$$

*Proof of Proposition 7.* Fix  $\epsilon > 0$  and let  $\rho \in (0, 1)$ . Define  $\tilde{X} \triangleq (1 - \rho)X$  and  $\tilde{Y} \triangleq \rho Y$ . We have that  $\tilde{X}$  and  $\tilde{Y}$  have exponential tails with parameters  $(c_X, \lambda_X/(1 - \rho))$  and  $(c_Y, \lambda_Y/\rho)$  respectively. This is because

$$\begin{aligned} 1 &= \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{c_X \exp(-\lambda_X x)} = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x/(1 - \rho))}{c_X \exp(-\lambda_X(x/(1 - \rho)))} = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(\tilde{X} > x)}{c_X \exp(-(\lambda_X/(1 - \rho))x)} \\ 1 &= \lim_{y \rightarrow \infty} \frac{\mathbb{P}(Y > y)}{c_Y \exp(-\lambda_Y y)} = \lim_{y \rightarrow \infty} \frac{\mathbb{P}(Y > y/\rho)}{c_Y \exp(-\lambda_Y(y/\rho))} = \lim_{y \rightarrow \infty} \frac{\mathbb{P}(\tilde{Y} > y)}{c_Y \exp(-(\lambda_Y/\rho)y)} \end{aligned}$$

Since  $\tilde{X}$  and  $\tilde{Y}$  have an exponential tail with parameters  $(c_X, \lambda_X/(1 - \rho))$  and  $(c_Y, \lambda_Y/\rho)$  respectively, there exists constants  $\tilde{x}_0, \tilde{y}_0$  such that for all  $t \geq t_0 = \max\{\tilde{x}_0, \tilde{y}_0\}$ , we have that

$$(1 - \epsilon)c_X \exp(-(\lambda_X/(1 - \rho))t) \leq \mathbb{P}(X > t) \leq (1 + \epsilon)c_X \exp(-(\lambda_X/(1 - \rho))t) \quad (\text{D.16})$$

$$(1 - \epsilon)c_Y \exp(-(\lambda_Y/\rho)t) \leq \mathbb{P}(Y > t) \leq (1 + \epsilon)c_Y \exp(-(\lambda_Y/\rho)t) \quad (\text{D.17})$$

Define  $\lambda_{\tilde{X}} \triangleq \lambda_X/(1 - \rho)$  and  $\lambda_{\tilde{Y}} \triangleq \lambda_Y/\rho$ . Furthermore, we define  $\underline{\lambda} \triangleq \min\{\lambda_{\tilde{X}}, \lambda_{\tilde{Y}}\}$  and  $\bar{\lambda} \triangleq \max\{\lambda_{\tilde{X}}, \lambda_{\tilde{Y}}\}$ .

**Upper Bound on  $\mathbb{P}(Z > t)$**  Next we want to provide an upper bound on the tail of  $Z$ . Choose an  $s \in (0, \underline{\lambda})$ . We will optimize for  $s$  later. Then we have that for  $t > 1$ ,

$$\mathbb{P}(Z > t) \stackrel{(a)}{=} \mathbb{P}(\exp(sZ) > \exp(st)) \stackrel{(b)}{\leq} \mathbb{E}[\exp(sZ)] \cdot \exp(-st) \stackrel{(c)}{=} \mathbb{E}[\exp(s\tilde{X})]\mathbb{E}[\exp(s\tilde{Y})] \exp(-st),$$

where (a) follows from the fact that  $\exp(sx)$  is strictly increasing, (b) follows from Markov's inequality and (c) follows from the fact that  $Z = \tilde{X} + \tilde{Y}$  and  $\tilde{X}$  and  $\tilde{Y}$  are independent.

We will now show that there exists constants  $c'_X = c(\epsilon, t_0, c_X)$  and  $c'_Y = c(\epsilon, t_0, c_Y)$  such that

$$\mathbb{E}[\exp(s\tilde{X})] \leq c'_X \left(1 - \frac{s}{\lambda_{\tilde{X}}}\right)^{-1}, \quad \mathbb{E}[\exp(s\tilde{Y})] \leq c'_Y \left(1 - \frac{s}{\lambda_{\tilde{Y}}}\right)^{-1}. \quad (\text{D.18})$$

We will show this for  $\mathbb{E}[\exp(s\tilde{X})]$  and the steps for  $\mathbb{E}[\exp(s\tilde{Y})]$  will follow analogously. Further, we will assume that  $t_0 > \max\{1, (1 + \epsilon)c_X\}$ . If  $t_0 < \max\{1, (1 + \epsilon)c_X\}$ , the analysis will require trivial modifications. We have that

$$\begin{aligned} \mathbb{E}[\exp(s\tilde{X})] &\stackrel{(a)}{=} \int_0^\infty \mathbb{P}(\exp(s\tilde{X}) > t) dt, \\ &\stackrel{(b)}{=} 1 + \int_1^\infty \mathbb{P}(\tilde{X} > \ln t/s) dt, \\ &\stackrel{(c)}{\leq} 1 + \int_1^{t_0} 1 dt + \int_{t_0}^\infty (1 + \epsilon)c_X \exp(-(\lambda_X/s(1 - \rho)) \ln t) dt, \\ &\stackrel{(d)}{\leq} t_0 + (1 + \epsilon)c_X \int_1^\infty \exp(-(\lambda_X/s(1 - \rho)) \ln t) dt, \\ &\stackrel{(e)}{=} t_0 + (1 + \epsilon)c_X \frac{1}{(\lambda_X/(s(1 - \rho))) - 1}, \\ &\stackrel{(f)}{\leq} \frac{c'}{(\lambda_X/(s(1 - \rho))) - 1}, \end{aligned}$$

where (a) follows from tail sum formula for expectation [164], (b) follows from the fact that  $\mathbb{P}(\exp(s\tilde{X}) > t) = \mathbb{P}(\tilde{X} > \ln t/s)$ , (c) follows from D.16, (d) follows trivially, (e) follows from the fact that  $\int_1^\infty \exp(-(\lambda_X/s(1 - \rho)) \ln t) dt = \frac{s(1-\rho)}{\lambda_X - s(1-\rho)}$ , (f) follows from an appropriate choice of  $c'$ .

Therefore, we have that

$$\begin{aligned} \mathbb{P}(Z > t) &\stackrel{(a)}{\leq} c'_X c'_Y \exp\left(-st - \ln\left(1 - \frac{s}{\lambda_{\tilde{X}}}\right) - \ln\left(1 - \frac{s}{\lambda_{\tilde{Y}}}\right)\right), \\ &\stackrel{(b)}{\leq} c'_X c'_Y \exp\left(-st - \left(\frac{\lambda}{\lambda_{\tilde{X}}} + \frac{\lambda}{\lambda_{\tilde{Y}}}\right) \ln\left(1 - \frac{s}{\lambda}\right)\right), \\ &\stackrel{(c)}{=} \exp(\lambda) c'_X c'_Y t^{1+(\lambda/\lambda)} \exp(-\lambda t), \end{aligned}$$

where (a) follows from (D.18), (b) follows from the fact that  $-\ln\left(1 - \frac{s}{\lambda_X}\right) \leq -\frac{\lambda}{\lambda_X} \ln\left(1 - \frac{s}{\lambda}\right)$

and  $-\ln\left(1 - \frac{s}{\lambda_Y}\right) \leq -\frac{\lambda}{\lambda_Y} \ln\left(1 - \frac{s}{\lambda}\right)$  and (c) follows from choosing  $s = (1 - 1/t)\lambda$ . Define  $C = \exp(\lambda)c'_X c'_Y$ . Then we have that  $\mathbb{P}(Z > t) \leq Ct^{1+(\bar{\lambda}/\lambda)} \exp(-\lambda t)$ .

**Lower Bound on  $\mathbb{P}(Z > t)$**  Next we want to provide a lower bound on the tail of  $Z$ . Define  $L = \tilde{X}\mathbb{1}\{\lambda_{\tilde{X}} < \lambda_{\tilde{Y}}\} + \tilde{Y}\mathbb{1}\{\lambda_{\tilde{X}} > \lambda_{\tilde{Y}}\}$ . Let  $\underline{c} = \min\{c_X, c_Y\}$ . For all  $t \geq t_0$ , we have that,

$$\mathbb{P}(Z > t) \stackrel{(a)}{\geq} \mathbb{P}(L > t) \stackrel{(b)}{\geq} (1 - \epsilon)\underline{c} \exp(-\lambda t) := c \exp(-\lambda t)$$

Using D.2, for all  $t \geq t_0$ , we have that

$$1 - (1 - c \exp(-\lambda t))^n \leq \mathbb{P}(Z_{(n:n)} > t) \leq 1 - (1 - Ct^{1+(\bar{\lambda}/\lambda)} \exp(-\lambda t))^n \quad (\text{D.19})$$

Choose a  $\delta \in (0, \lambda)$ . There exists  $t'_0$  such for all  $t \geq t'_0$ , we have that  $t^{1+\bar{\lambda}/\lambda} \leq \exp(\delta t)$ . Therefore, we have that

$$1 - (1 - c \exp(-\lambda t))^n \leq \mathbb{P}(Z_{(n:n)} > t) \leq 1 - (1 - C \exp(-(\lambda - \delta)t))^n \quad (\text{D.20})$$

Using the analysis in the proof of Proposition 6, we have that

$$\frac{1}{\lambda} \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{\ln n} \leq \limsup_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{\ln n} \leq \frac{1}{\lambda - \delta}$$

Since this set of inequalities is true for all  $\delta \in (0, \lambda)$ , we have that  $\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{(n:n)}]}{\ln n} = \frac{1}{\lambda} = \max\{\lambda_X/(1 - \rho), \lambda_Y/\rho\}$ . This completes the proof. ■

### D.2.3 Proof of Theorem 10

*Proof of Theorem (10.a).* Recall from the proof of Theorem (9.a), we have that  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n) = (1 - \rho)\mathbb{E}[q_{(n:n)}] - (1 - \rho)\mu_q$ . Now using Proposition 6, the result follows. ■

*Proof of Theorem (10.b).* Define  $Z_k = (1 - \rho)q_k + \rho\varphi_k$ . We have that

$$\frac{\Delta_{q \rightarrow u}^{\text{uncap}}(n)}{\ln n} = \frac{\mathbb{E}[Z_{(n:n)}]}{\ln n} - \frac{(1 - \rho)\mathbb{E}[q_{(n:n)}]}{\ln n} - \frac{\rho\mu_\varphi}{\ln n}$$

Now using Propositions 6 and 7, the result follows. ■

#### D.2.4 Proof of Theorem 12

*Proof of Theorem (12.a).* The proof of Theorem (12.a) mimics the proof of Theorem (11.a) and hence is omitted. ■

*Proof of Theorem (12.b).* The proof of Theorem (12.b) will make use of Lemma 4 and Proposition 6. Let us denote  $\varphi_{k,(n-k:n-k)} := \varphi_{(n-k:n-k)}$ . Fix  $\epsilon > 0$ . There exists an  $k_0 \in \mathbb{N}$  such that for all  $k \geq k_0$ , we have that

$$(1 - \epsilon) \ln k / \lambda \leq \mathbb{E}[\varphi_{(k:k)}] \leq (1 + \epsilon) \ln k / \lambda \tag{D.21}$$

Recall that  $\Phi_n \triangleq n^{-1} \sum_{k=1}^n \mathbb{E}[\varphi_{(k:k)}]$ . We can upper bound  $\Phi_n$  as follows:

$$\begin{aligned} \Phi_n &\stackrel{(a)}{=} \frac{1}{n} \sum_{k=1}^{k_0} \mathbb{E}[\varphi_{(k:k)}] + \frac{1}{n} \sum_{k=k_0}^n \mathbb{E}[\varphi_{(k:k)}], \\ &\stackrel{(b)}{\leq} \mu_\varphi \frac{k_0(k_0 + 1)}{2n} + \frac{1}{n} (1 + \epsilon) / \lambda \sum_{k=k_0}^n \ln k, \\ &\stackrel{(c)}{\leq} \mu_\varphi \frac{k_0(k_0 + 1)}{2n} + \frac{1}{n} (1 + \epsilon) / \lambda \int_1^n \ln x \, dx, \\ &\stackrel{(d)}{\leq} \mu_\varphi \frac{k_0(k_0 + 1)}{2n} + (1 + \epsilon) \ln n / \lambda, \end{aligned}$$

where (a) follows trivially, (b) follows from (D.21) and the fact that  $\mathbb{E}[\varphi_{(k:k)}] \leq k\mu_\varphi$  for all  $k \leq k_0$  since  $\mathbb{E}[\max\{X_1, X_2, \dots, X_k\}] \leq \mathbb{E}\left[\sum_{j=1}^k X_j\right] = k\mu_\varphi$ , (c) follows trivially and (d) follows

from the fact that  $\int \ln x dx = x \ln x - x$ . Using this, we have that

$$\limsup_{n \rightarrow \infty} \frac{\Phi_n}{\ln n / \lambda} \leq 1 + \epsilon.$$

Using similar arguments as above, we can easily show that

$$\liminf_{n \rightarrow \infty} \frac{\Phi_n}{\ln n / \lambda} \geq 1 - \epsilon.$$

Since this holds for all  $\epsilon > 0$ , combining the two, we have that  $\lim_{n \rightarrow \infty} \frac{\Phi_n}{\ln n / \lambda} = 1$  and this completes the proof. ■

### D.3 (Partial) Results for bounded utility distributions

In this section, we discuss the case where the common and the idiosyncratic terms are drawn from bounded distributions  $P_q$  and  $P_\varphi$ . For simplicity we will assume that both the distributions  $P_q$  and  $P_\varphi$  are continuous distributions with density bounded below and above over the interval  $[a, b]$ , where  $a \geq 0$  and  $b < \infty$ .

**Theorem 14 (Uncapacitated Supply, Bounded Distribution)** *Consider the uncapacitated supply setting. Assume that the common terms  $(q_y)$  are drawn i.i.d from a continuous distribution  $P_q$  with support  $[a, b]$ , finite mean  $\mu_q < \infty$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from a continuous distribution  $P_\varphi$  with support  $[a, b]$ , finite mean  $\mu_\varphi < \infty$ . For any  $\rho \in [0, 1]$ , we have that,*

(14.a) *The difference in the agent welfare  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  obtained in the Only Quality Information regime and the No Information regime is given as*

$$\lim_{n \rightarrow \infty} \Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n) = (1 - \rho) \cdot (b - \mu_q).$$

(14.b) The difference in the agent welfare  $\Delta_{\emptyset \rightarrow \varphi}^{\text{uncap}}(n)$  obtained in the Full Information regime and Only Quality Information regime is

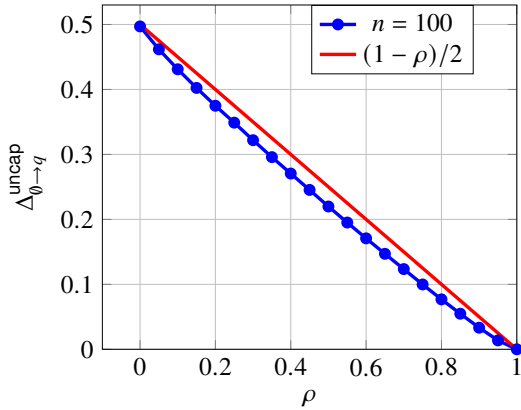
$$\lim_{n \rightarrow \infty} \Delta_{q \rightarrow u}^{\text{uncap}}(n) = \rho \cdot (b - \mu_\varphi).$$

*Proof of Theorem (14.a).* Recall from proof of Theorem (9.a), we have that  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n) = (1 - \rho)\mathbb{E}[q_{(n:n)}] - (1 - \rho)\mu_q$ . We have that  $\lim_{n \rightarrow \infty} \mathbb{E}[q_{(n:n)}] = b$ . This completes the proof. ■

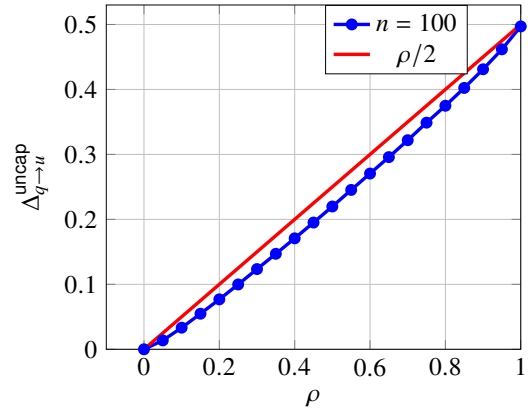
*Proof of Theorem (14.b).* Define  $Z_k = (1 - \rho)q_k + \rho\varphi_k$ . We have that

$$\Delta_{q \rightarrow u}^{\text{uncap}}(n) = \mathbb{E}[Z_{(n:n)}] - (1 - \rho)\mathbb{E}[q_{(n:n)}] - \rho\mu_\varphi.$$

We have that  $\lim_{n \rightarrow \infty} \mathbb{E}[Z_{(n:n)}] = \lim_{n \rightarrow \infty} \mathbb{E}[q_{(n:n)}] = b$ . This completes the proof. ■



(a) Theorem (14.a)



(b) Theorem (14.b)

Figure D.1: Simulations plot of  $\Delta_{\emptyset \rightarrow q}^{\text{uncap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{uncap}}(n)$  as a function of  $\rho \in [0, 1]$  when  $P_q$  and  $P_\varphi$  are the Uniform( $[0, 1]$ ).

**Theorem 15 (Capacitated Supply, Bounded Distribution)** Consider the capacitated supply setting. Assume that the common terms  $(q_y)$  are drawn i.i.d from a continuous distribution  $P_q$  with support  $[a, b]$ , finite mean  $\mu_q < \infty$ . Assume that the idiosyncratic terms  $(\varphi_{xy})$  are drawn i.i.d from a continuous distribution  $P_\varphi$  with support  $[a, b]$ , finite mean  $\mu_\varphi < \infty$ . For any  $\rho \in [0, 1]$ , we have that,

(15.a) *The difference in the agent welfare  $\Delta_{\emptyset \rightarrow q}^{\text{cap}}(n)$  obtained in the Only Quality Information regime and the No Information regime is given as*

$$\lim_{n \rightarrow \infty} \Delta_{\emptyset \rightarrow q}^{\text{cap}}(n) = 0.$$

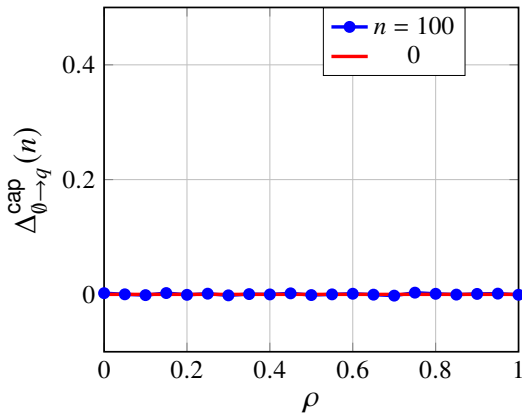
(15.b) *The difference in the agent welfare  $\Delta_{\emptyset \rightarrow \varphi}^{\text{cap}}(n)$  obtained in the Full Information regime and Only Quality Information regime is*

$$\rho b - ((1 - \rho)\mu_q + \rho\mu_\varphi) \leq \lim_{n \rightarrow \infty} \Delta_{q \rightarrow u}^{\text{cap}}(n) \leq \rho \cdot (b - \mu_\varphi).$$

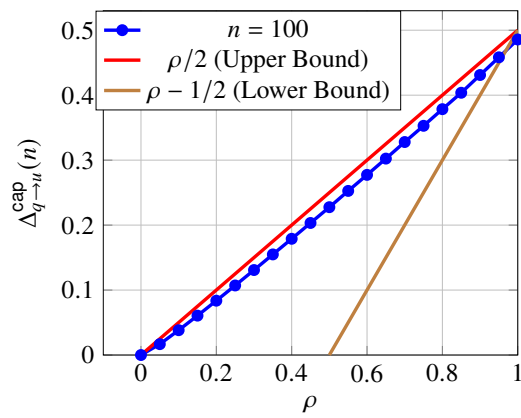
*Proof of Theorem (15.a).* The proof follows the same argument as the proof of Theorem (11.a). ■

*Proof of Theorem (15.b).* The proof follows from Lemma 4 and the fact that  $\lim_{n \rightarrow \infty} \Phi_n = b$ . ■

**Remark 9** *In Theorem (15.b), we provide an upper and lower bound on the asymptotic marginal welfare gain of personalizing recommendations. From Figure D.2b, we observe that there is a gap between the upper and lower bounds for the case of bounded distribution. In general, it is a challenging problem to provide a crisp characterization for the marginal welfare gains of personalizing recommendations and as such we defer this question for future research.*



(a) Theorem (15.a)



(b) Theorem (15.b)

Figure D.2: Simulations plot of  $\Delta_{0 \rightarrow q}^{\text{cap}}(n)$  and  $\Delta_{q \rightarrow u}^{\text{cap}}(n)$  as a function of  $\rho \in [0, 1]$  when  $P_q$  and  $P_\varphi$  are the Uniform( $[0, 1]$ ).