Study of Customer Behavior in a Revenue Management Setting Using Data-Driven Approaches

Sareh Nabi-Abdolyousefi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Hamed Mamani, Chair

Mark Hillier

Foad Iravani

Program Authorized to Offer Degree:
Business Administration

University of Washington

Abstract

Study of Customer Behavior in a Revenue Management Setting Using Data-Driven Approaches

Sareh Nabi-Abdolyousefi

Chair of the Supervisory Committee:

Professor Hamed Mamani

Information System and Operations Management

The objective of this study is to propose novel dynamic pricing mechanisms in the presence of strategic customers using data-driven approaches. Dynamic pricing is the latest trend in pricing strategies and allows optimal response to real-time demand and supply information. Firms often face uncertainties when making pricing decisions. One of the uncertainties often involved is unknown demand. Therefore, businesses seek to optimize revenue while learning demand and reducing the uncertainty involved in setting prices. Understanding consumer decision-making is another crucial aspect of pricing in revenue management. One of the detrimental effects of dynamic pricing is that it invokes a type of behavior in customers that is referred to as forward-looking, or strategic, in revenue management literature. The strategic customer considers future price decreases, and purchases the product if his or her discounted surplus is higher than the immediate surplus.

In chapters 1 and 2, we study a retailer who is pricing dynamically to maximize his expected cumulative revenue. We assume that the retailer has no information regarding expected demand nor the type of customers he is facing, whether they are myopic or strategic in

their shopping behavior. In the problem of dynamic pricing under demand uncertainty, we face an inherent trade-off between the exploration involved in learning demand and the exploitation which occurs due to revenue maximization. One way of modeling this trade off is using the multi-arm bandit modeling approach. Many algorithms have been proposed to solve stochastic multi-arm bandit problems. Our focus is on the Thompson Sampling (TS) algorithm which takes a Bayesian approach and was introduced by William R. Thompson. We propose a pricing mechanism called Strategic Thompson Sampling algorithm which is built upon the TS algorithm. Our main contribution in these two chapters is to merge the literature on strategic behavior with the literature on dynamic pricing and demand learning based on the classical multi-arm bandit modeling approach. In these chapters, the retailer is applying our proposed Strategic Thompson Sampling algorithm to learn expected demand in an exploration-versus-exploitation fashion.

We start our analysis with a Bernoulli demand scenario in chapter 1 and extend our work to a Normal demand scenario in chapter 2. For both Bernoulli and Normal demand scenarios, we demonstrate numerically that the retailer's long run price offer decreases as the patience level of the strategic customer increases. We further show that the retailer can be better off in terms of his expected cumulative revenue when facing strategic customers. One potential explanation for this observation is the retailer's lower exploration of non-optimal arms in the presence of strategic customers rather than myopic ones. Our intuition is analytically and numerically confirmed for both Bernoulli and Normal demand scenarios. We further provide and compare expected regret bounds on the retailer's expected cumulative revenue for both types of customers. We conclude that the retailer's regret is lower when facing strategic customers as compared to myopic ones.

Our objective in chapter 3 is to improve our starting point by building an informative prior and more specifically, an empirical Bayes prior for the Bayesian online learning algorithm that performs binary prediction. The underlying model used in this chapter is a Bayesian

Linear Probit (BLIP) model which performs binary classification on a public data set called "Census Income Data Set". Our goal is to build an informative prior using a portion of the training data set and start the BLIP model with the built-in prior rather than the non-informative standard Normal distributions. We further compare the prediction accuracies of the BLIP model with informative and non-informative priors. An empirical Bayes model (Blip with empirical Bayes prior) has been implemented recently in the production system of one of the largest online retailers. The web-lab experiment is currently running.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My deepest gratitude goes to my family for all the love, inspiration, support, and encouragement they have given me. Whatever I do in my life is a very small token of appreciation to my parents who put their heart and soul to raise my siblings and me. They cultivated life lessons in us by living them themselves: by being wise, inspirational, responsible, hard-working, supportive, kind, respectful; loving; and having good morals. My heartfelt thanks to my older sister, Marzieh, who played a remarkable role in the personal and career growth of my siblings and me. She is an angel in our lives and we are forever grateful to her.

My deepest gratitude and sincere thanks to my supervisor, Prof. Hamed Mamani, for all his support and encouragement during my PhD program. He is the best supervisor I could ever ask for. He has always given me the freedom to follow my passion and has supported me fully. I'm also very grateful for the great insights, knowledge, and intelligence he brought to this body of work.

I would also like to thank my reading committee and committee members, Prof. Mark Hillier, Prof. Archis Ghate, Prof. Thomas Richardson, and Prof. Foad Iravani for their time, support, and insightful feedback along the way.

I would like to express my deep gratitude to Prof. Sham Kakade who provided the opportunity for me to work on the third chapter of this body of work. And huge thanks to Houssam Nassif and Prof. Guido Imbens whose tremendous support and help along the way made this chapter an impactful one. I would also like to express my profound gratitude to Prof. Kevin Jamieson for his tremendous help in deriving the analytical results of chapter 2. His great

support, insights, and intelligence made the derivation more clear after our discussions.

There were three courses at UW that greatly contributed to my graduate education: Econometrics from Prof. Thomas Richardson; Machine Learning from Prof. Sham Kakade; and Online and Adaptive Methods for Machine Learning from Prof. Kevin Jamieson. These courses reinforced my interest in this field and equipped me with the knowledge I needed to finish this body of work, enabling me to pursue my academic and professional passions after graduation. My sincere thanks to these wonderful and knowledgeable teachers.

I would also like to thank Foster administrators, Shawna Reimers, Jessica Aceves, and Jaime Banaag, who went above and beyond to help whenever needed. My sincere thanks, also, to Prof. Yong-Pin Zhou, Prof. Ted Klastorin, and Prof. Ed Rice, Meera Mahabala, Milinda Vitharana, and TJ Vassar for all the support along the way.

Finally, I would like to thank all my friends who made the PhD program more enjoyable and fun: Elnaz Jalilipour Alishah, Behnaz Ghahestani Bojd, Aravinda Garimella, Pegah Jalali, Shahryar Doosti, Emisa Nategh, Amir Fazli, Omid Rafieian, Mohammad Ebrahim Arbabian, Soraya Fatehi, Mina Ekramnia, Eugene Pavlov, Michelle Lucena, and Prof. Shima Nassiri. I would also like to extend a "heartful thank you" to my close friends who have always supported me and visited me from far away: Roya Safaei, Fatemeh Arbab, and Mahjabeen Ahmed.

DEDICATION

To Goodarz and Nosrat (Dad & Mom)

&

To my siblings: Marzieh, Aboozar, Soheila, Razieh and Negin

Chapter 1

DYNAMIC PRICING FOR STRATEGIC CUSTOMERS: BERNOULLI DEMAND

1.1 Introduction

Pricing goods and services is the fastest and most effective way to influence the profitability of any business. According to an article in Harvard Business Review [64], a 1% increase in price generates an 11.1% increase in profit on average assuming demand remains constant. According to a recent article on Due.com [57], pricing too low can hurt brand reputation and the perceived quality of products and services. On the other hand, pricing that is too high can also be detrimental to revenue in the presence of a competitor with a similar, but cheaper, product. Obtaining the right pricing strategy has therefore been the focus of researchers and practitioners from the beginning of the revenue management era. Nowadays, more and more businesses apply revenue management optimization tools to determine optimal price. For extensive studies on pricing strategies in the revenue management literature, refer to [70] and [82].

Dynamic pricing is the latest trend in pricing strategies and allows optimal response to real-time demand and supply information. According to an article in The Economist [75] in January 2016, the practice of dynamic pricing dates back to the 1980s, when American Airlines varied prices to compete with low cost airlines of the time such as People's Express. Other airlines, hotels, and car rental firms followed suit with price variation in the 1990s. [39] reported that 90% of the 32 largest online retailers perform price experimentation. Nowadays, more firms are practicing dynamic pricing. In recent years, according to an article in The

Wall Street Journal [67] in 2015, more businesses such as Uber, Lyft, and SeaWorld are adjusting prices based on demand realizations.

Dynamic pricing has given flexibility and new capabilities to revenue management. Especially with the advent of e-commerce, it has become easier to track customer decisions, monitor demand, and adjust prices dynamically at negligible costs. According to a recent report released by the U.S. Commerce Department [68], e-commerce sales in 2016 were estimated at $\$394.9B$. Total e-commerce sales experienced a growth rate of $15.1\%(\pm 1.8\%)$ in 2016, compared to the previous year's $\$341.70B$ sales estimate. Amazon, Inc., as one of the global e-commerce giants, makes up a high portion of these online sales by adjusting prices frequently, sometimes in a matter of minutes. To perform its pricing changes, Amazon incorporates customers' purchase behavior, income level, and geographic information. According to an article in Business.com [49], Walmart changes its product prices about 50000 times a month. By applying dynamic pricing, Walmart was able to increase global online sales by 30% in 2013.

Firms often face uncertainties when making pricing decisions. One of the uncertainties often involved is unknown demand. Therefore, businesses seek to optimize revenue while learning demand and reducing the uncertainty involved in setting prices. This is known as the problem of "learning and earning" and goes back to the pioneering works by economists such as [72]. The two primary approaches for learning demand in the economics literature are model-based and data-driven. In model-based demand learning, a particular functional form for demand is assumed which has finite unknown parameters. The second approach is data-driven where demand is learned solely from the customer's prior purchase history. For comprehensive surveys on dynamic pricing under uncertainty refer to [28], Cesa-Bianchi (2012), and [11].

In this work, we focus on the data-driven approach. In the problem of dynamic pricing under demand uncertainty, we face an inherent trade off between the exploration involved in learning demand and the exploitation which occurs due to revenue maximization. One way

of modeling this trade-off is using the multi-armed bandit approach which is the simplest form of reinforcement learning. The multi-armed bandit (MAB) modeling approach was originally proposed by [71] and has received extensive attention in the operations management literature. Many algorithms have been proposed to solve stochastic MAB problems ([40], [41], [7], [38], [46], [19, chap. 2], ([84]), [47], [3], [47]).

The performance of algorithms proposed for solving MAB problems are measured in terms of "regret". Regret measure goes back to the work by [14]. Regret in each period is defined as the expected loss due to not playing the optimal arm. Regarding theoretical guarantees, [51] provided an optimal asymptotic lower bound on the expected regret of any bandit algorithm. Specifically, they showed that as T goes to infinity, expected regret can be of order $O(\ln T)$ and no bandit algorithm can have a better asymptotic performance. This asymptotically optimal lower bound has been a reference point for comparing theoretical guarantees of proposed bandit algorithms thereafter.

Our focus is on the Thompson Sampling (TS) algorithm, also known as posterior sampling, which dates back to 1933 when it was proposed by William R. Thompson ([84]). The Thompson Sampling algorithm is a member of the Probability Matching family of algorithms which takes a Bayesian approach to solve stochastic multi-arm bandit problems. In each round, the algorithm selects an arm based on its posterior probability of being the optimal arm. In traditional Thompson sampling, the reward of each arm follows a Bernoulli distribution where the expected reward is unknown to the agent. The agent's objective is to find the optimal arm that gives the maximum expected cumulative reward. Compared to state-of-the-art algorithms, Thompson Sampling has demonstrated promising empirical results ([47], [23], [76]) and significant theoretical guarantees ([60], [47], [2], [3]).

Understanding consumer decision-making is another crucial aspect of pricing in revenue management. According to an article by [79], about 50% of customers selected price as among the top three main drivers of their purchase decisions, and 18% of customers chose it as the most influential factor. This study included approximately 30000 customers across a variety

of industries in the United States including consumer products, information technology (IT), healthcare, apparel and financial services. Nowadays, with the quickly growing rate of online shopping, customers can track prices and detect patterns in price variations easily. One of the detrimental effects of dynamic pricing is that it invokes a type of behavior in customers that is referred to as forward-looking, or strategic, in revenue management literature. The strategic customer considers potential future price decreases, and purchases the product if his or her discounted surplus is higher than the immediate surplus. The opposite behavior, in which the customer purchases a product as soon as the price drops below his or her valuation of the product, is referred to as myopic behavior.

Strategic behavior was first informally introduced by [25] conjecture and later formalized by [80] and [20]. The Coase conjecture states that a monopolist selling a durable good can not charge a price higher than the marginal cost in the presence of infinite patient customers. The existence and prevalence of strategic behavior has been confirmed through recent empirical studies ([69], [55], [59], [78], etc). For example, [55] built a structural model to estimate the fraction of strategic customers in the travel industry. They found that 5.2% to 19.2% of the population exhibits strategic behavior. [81] reports that strategic behavior poses a greater challenge for companies that practice dynamic pricing rather than applying ad-hoc strategies to determine their prices. He argues that strategic customers can easily detect regularities enforced by dynamic pricing, such as markups or markdowns.

It is a common belief that strategic behavior may negatively affect a firm's revenue. However, research demonstrates that misidentifying customers who shop strategically can actually lead to significant revenue loss. [10] reported revenue loss of about 20% if a strategic customer is misidentified as a myopic one. Moreover, [15] and [53] reported an expected revenue loss of more than 50% and 20% respectively for mistakenly identifying the strategic customer. [55] conclude that strategic behavior that exists in the travel industry, against the common belief, does not necessarily hurt the revenue and actually varies across the market. The idea that strategic behavior can have either a positive or negative impact on revenue has

been confirmed theoretically as well (e.g. [81] and [24]). The recent work by [1] sheds new light on the effects of strategic behavior, not only on firms – which has been the focus of operations management literature – but also on customers themselves and on the society. By endogenizing strategic behavior and considering the heterogeneous population of customers, [1] found that only high-valued customers benefit in the presence of strategic behavior; low-valued ones are indifferent, while moderate-valued customers are actually worse off. They also report that social welfare is always higher in the presence of myopic rather than forward-looking behavior.

A growing number of works in revenue management take the demand learning approach and perform dynamic pricing in the presence of strategic customers using a variety of rich models. [54] built an empirical demand model using an adaptation of the Aggregating Algorithm (AA) solely based on history of sales data. [35] built a multi-product demand prediction model based on a non-parametric machine learning algorithm (regression trees) for first-exposure styles of Rue La La, an online fashion retailer. The researchers fed the predicted demand to a multi-product static price optimization model to maximize revenue and reported an approximately 10% increase in revenue on the test group based on their field experiment. Some studies take a dynamic programming approach ([5], [34], [9], etc), or a mechanism design approach ([32], [37]). Closest to our work are the studies that take the multi-arm bandit modeling approach in order to learn demand ([13], and [36]).

While consumer behavior is an important issue, we have not, to the best of our knowledge, seen works that merge strategic behavior, dynamic pricing, and demand learning using the multi-arm bandit modeling approach. This motivated us to propose the Strategic Thompson Sampling algorithm to address this shortcoming in the literature. The price offer mechanism in our Strategic Thompson Sampling algorithm is close to the one described in [3]. A fundamental difference between their algorithm and ours is that we added strategic customer purchase decisions to the algorithm. Another recent work similar to ours in terms of demand learning approach is the work by [36]. They design and implement an algorithm based on

Thompson Sampling for a network revenue management setting. The algorithm samples demand from its posterior distribution and then solves a linear programming (LP) optimization problem to incorporate inventory constraints and obtain optimal prices. Even though the demand learning approach in our work and theirs is similar, as both utilize Thompson Sampling, the parameters of the two studies are quite different. They assumed finite inventory and no strategic behavior on the customer side; however, in our work, the inventory is unlimited and the presence of strategic behavior is one of the driving factors in determining optimal price.

The main contribution of chapters 1 and 2 is to merge the literature on strategic behavior with the literature on dynamic pricing and demand learning based on the classical multi-arm bandit modeling approach. We demonstrate numerically that the retailer's price offer in the long run decreases as the patience level of the customer increases. This result is compatible with similar scenarios in the revenue management literature ([15]). We further compare the retailer's expected cumulative revenue in the presence of strategic and myopic customers. Through an extensive parameter domain exploration, we found that the retailer can be better off in the presence of strategic behavior. This is due to less exploration of non-optimal arms when facing strategic customers. Finally, we derived upper bounds on the expected regret for the retailer's expected cumulative revenue for both strategic and myopic customers. We conclude that the retailer's regret is lower on average when facing strategic customers as compared to myopic ones.

1.2 Literature Review

The three streams of research mainly related to our analysis are (1) dynamic pricing and demand learning; (2) classical multi-armed bandit modeling approach for dynamic learning; and (3) strategic consumer behavior. There exists rich literature in each of these aforementioned areas which we will briefly detail here.

The first stream of research relevant to our work is demand learning in a dynamic pricing setting. Some studies in the literature assume that demand follows a functional form with unknown parameters and estimate the parameters using classical statistical methods such as maximum likelihood estimation (e.x [17], [48], [30], [18], [45], [44]). Some studies also take a parametric Bayesian approach to learn unknown demand parameters and maximize revenue ([45], [5],[26]). As another approach, some works take a non-parametric path to learn demand when a functional form can not be assumed ([16], [33]). For an overview of papers in this domain refer to [6] and [29].

The second stream of research related to our work is the classical multi-arm bandit modeling approach originally proposed by Robbins (1952). Since then, many algorithms and heuristics have been introduced to solve MAB problems. One of the initial approaches introduced are index allocation strategies originally proposed by [40]. Gittins index and potentially other index policies face an incomplete learning issue ([41]). One of the popular ad-hoc strategies introduced to solve stochastic multi-armed bandit problems are the Upper Confidence Bound (UCB) family of algorithms, which take a frequentist approach to find the optimal arm. Essentially, these algorithms compute the upper confidence bounds on the unknown parameters and select arms with the higher upper bounds ([7], [38], [46], and [58]). For an overview of the UCB family of algorithms, refer to [19, chap. 2].

The Probability Matching family of algorithms, on the other hand, take a Bayesian approach to tackle stochastic multi-arm bandit problems. The Thompson Sampling algorithm ([84]), as a member of this family of algorithms, has demonstrated promising empirical results. [76] reveals that Thompson Sampling performs reasonably well compared to the state of the art algorithms. [23] show empirically that expected regret for Thompson Sampling is comparable to the lower bound of [51]. Moreover, they showed in their experiments that Thompson Sampling is more robust to delayed or batched feedback. For the first time, [47] provide numerical experiments that demonstrate Thompson Sampling outperforms optimal policies introduced in the literature. Specifically, they compare Thompson Sampling with

the best version of the UCB family of algorithms and show that Thompson Sampling has the lowest expected regret in the long run (see [43], [42], and [61] for more references). Thompson Sampling has received considerable attention in industry as well (e.g. [76], [42], and [83]).

While the Thompson Sampling algorithm is easy to implement and numerically attractive, significant theoretical guarantees have been proposed recently. [60] showed asymptotic convergence of Thompson Sampling. [47] provided asymptotic optimality of Thompson Sampling. The work by [2] proves a logarithmic expected regret bound for this algorithm. [3] also introduced a novel martingale-based technique to perform regret analysis. More precisely, they provided optimal problem-dependent and near optimal problem-independent regret bounds on the Thompson Sampling algorithm with Beta and Gaussian priors. [47] refined the analysis of Thompson Sampling and presented its optimality for Bernoulli bandits. Moreover, [50] extended this algorithm to one-dimensional exponential family distributions. [73] provided an information based analysis of Thompson Sampling and obtained expected regret bounds using Bayesian risk.

The third stream of research that relates to our analysis is the impact of customers' strategic behavior in operations management. Rich literature has studied firms' optimal decisions on issues such as pricing ([12], [53], [52], [8], [32], [10], [81]), inventory ([56]), supply chain design ([21]) in the presence of strategic behavior. For an extensive review of these works refer to [66]. The presence of strategic behavior has been confirmed through empirical studies ([69], [55], [59], [78]). In the case of unlimited inventory, strategic customers only consider the likelihood of future price reduction in their purchase decision today ([15]). On the other side, with limited product supply, the strategic customer takes into account not only the price reduction, but also the likelihood of the product going out of stock ([27], [32], [10], [81]).

Strategic behavior has been studied in a variety of rich models. As one of the pioneer works, [15] characterize a subgame-perfect Nash equilibrium for a monopolist selling an unlimited inventory of a new product and demonstrate that a decreasing price strategy is optimal.

Assuming finite inventory, [81] study the interplay between firms' pricing strategy and a forward-looking customer's purchase decisions in a game theoretic model. He concludes that the heterogeneity in consumer population in both the valuation of the product and the patience level derives the optimal pricing strategy. [10]and [27] study a two-periods dynamic pricing problem of a seller facing finite inventory in the presence of strategic consumer behavior. [10] analyze two selling strategies, contingent-discounting and announced fixed-discount strategies and characterize a unique subgame-perfect Nash equilibrium consisting of threshold policies. [27] propose a class of pre-announced pricing policies and characterize the equilibrium in the game between the seller and strategic customers using a novel approach based on ordinary differential equations.

In contrast to other studies, [1] endogenize strategic behavior by adapting the model of "rational ignorance" from the economics literature (Downs 1957). They are mainly interested in understanding whether customers choose to behave strategically if given the choice. They report that any price or inventory commitment mechanism that reduces forward-looking behavior increases social welfare and benefits society as a whole. Moreover, effective selling strategies such as price commitment ([10], [62]) and inventory commitment ([81], [56], [53]) are introduced in the literature to increase firms' profit by reducing strategic waiting. For an overview of consumer choice and strategic behavior models in revenue management, refer to [77].

1.3 Model Description

In our setting, a retailer is selling an unlimited inventory of a single product to customers who behave strategically; i.e., customers who may wait to get a better deal in the next period. For simplicity, we refer to the retailer as a male, and to the customer as a female. The set of admissible prices that the retailer chooses from in each period is a finite discrete set denoted by $\{p_1, p_2, ..., p_N\}$. This convention is widely used in the literature, see [82]. Customer's patience level (valuation discounted factor) is denoted by δ which varies in $[0, 1]$.

$\delta = 1$ represents a pure strategic customer and $\delta = 0$ denotes a pure myopic one. Values of δ between 0 and 1 show a mixture of strategic and myopic behavior. As δ increases, the customer becomes more strategic and is willing to wait more in order to obtain a better deal.

We are modeling this scenario as a stochastic multi-armed bandit problem. There are N arms, indexed by $i = 1, ..., N$ and T periods. Price p_i is assigned to arm $i, \forall\ i \in \{1, 2, ..., N\}$. Pulling arm i is used interchangeably with selecting price p_i by the retailer. The true expected demand under price p_i is denoted by d_i for a myopic customer ($\delta = 0$), and $\tilde{d}_i$ for a strategic customer with patience level δ. The retailer does not know the expected demand associated with each price nor the type of customer he is facing in advance.

The stream of actions in each period are as follows. First, the retailer offers a price p_i among the admissible set of prices, $\{p_1, p_2, ..., p_N\}$. The retailer's price offer decision follows a Bayesian framework. He assumes a prior distribution on the unknown expected demand (or revenue) parameter. He then offers a price in each period that maximizes his expected empirical revenue based on sample draws from demand's (revenue's) posterior distributions. Then, a strategic customer with patience level δ observes the offered price and decides whether not to purchase, purchase immediately, or wait. The customer's decision is based on her valuation of the product and her utility comparison between the current and the next period. Then, the retailer observes the customer's purchase decision and updates his belief according to the expected demand (revenue) parameter associated with the price offered.

We start our analysis with the Bernoulli demand case explained in detail in Sections 1.3 to 1.6 and extend it to the Normal demand scenario in Chapter 2. We assume that the customer's valuation of the product under price p_i, v_i, follows a Bernoulli distribution, $v_i \sim Bernoulli(d_i)$. Myopic and strategic customers have the same valuation of the product. The only difference is that the strategic customer may delay her purchase. Furthermore, we assume that a population is present at the beginning of each period. This assumption is consistent with the work by [15]. The population is homogeneous in terms of customer's patience level, δ. The size of the population and prices are normalized to one. A summary

$N_{i,t}$	Number of times arm i is chosen up to time t
$R_{i,t}$	Number of success of arm i up to time t
d_i	Myopic customer's expected demand under price p_i
$\tilde{d}_i$	Strategic customer's expected demand under price p_i
v_i	Myopic/Strategic Customer's valuation
	under price p_i
r_t	Revenue observed at time t
$\bar{w}_{i,t}$	Weighted average of price p_i offers up to time t
p^{est}	Customer's estimation of next period price
δ	Customer's patience level (valuation discounted factor), $\delta \in [0,1]$

Table 1.1: Summary of notations

of our notations is presented in Table 1.1.

1.3.1 Expected Demand and Customer Type

In this subsection, we derive the relation between expected demand under price p_i for myopic customers (d_i) and strategic customers ($\tilde{d}_i$) for the Bernoulli demand scenario. Assuming price p_i is proposed by the retailer, the customer's purchase probability (expected demand) with patience level δ is given by

$$p_r(\text{purchase}) = p_r(\text{purchase}|v_i = 1)\ p_r(v_i = 1) + p_r(\text{purchase}|v_i = 0)\ p_r(v_i = 0) \qquad (1.1)$$

Since

$$p_r(\text{purchase}|v_i = 0) = 0, \quad p_r(v_i = 1) = d_i,$$

we obtain

$$\begin{aligned} p_r(\text{purchase}|v_i = 1) &= p_r(1 - p_i \geq \delta(1 - p^{est})) \\ &= p_r(p^{est} \geq 1 - \frac{1 - p_i}{\delta}). \end{aligned}$$

The first equality comes from the strategic behavior modeling assumption in the operations management literature. They assume that the strategic customers' purchase decision is based on the following utility maximization problem:

$$\max \{0, v_i - p_i, \delta(v_i - p^{est})\}. \tag{1.2}$$

In equation (1.2), v_i denotes the customer's valuation of the product under price p_i, δ represents the customer's patience level, and p^{est} is the customer's estimation of the next period price. Hence, by applying equations (1.1) and (1.2), we obtain the following relation between d_i and $\tilde{d}_i$:

$$\tilde{d}_i = d_i \ * \ p_r(p^{est} \geq 1 - \frac{1 - p_i}{\delta}). \tag{1.3}$$

The strategic customer with patience level δ computes the probability term in equation (1.3) based on the observed weighted average of prices. More specifically,

$$p_r(p^{est} \geq 1 - \frac{1 - p_i}{\delta}) = \sum_j \bar{w_j}, \qquad \forall j \ \ s.t. \ \ p_j \geq 1 - \frac{1 - p_i}{\delta}.$$

Here, $\bar{w_j}$ represents the current weighted average of price p_j offers. Note that the computation of the probability term is based on the revealed prices rather than the admissible set of prices from which the retailer can potentially offer. For the specific values of $\delta = 0$, and $\delta = 1$, expected demand is reduced to

$$\begin{aligned} &\text{If} \ \ \delta = 0 \ \ \rightarrow \ \ \tilde{d}_i = d_i \\ &\text{If} \ \ \delta = 1 \ \ \rightarrow \ \ \tilde{d}_i = d_i \ * \ p_r(p^{est} \geq p_i) \end{aligned}$$

Under the current proposed price p_i, the expected demand for a fully strategic customer ($\delta = 1$) is the myopic customer's expected demand (d_i) multiplied by the probability that the next period price is higher than the current offered price.

1.3.2 Modeling Customer Returns

As mentioned above, customers make purchase decisions based on their utility maximization given in (1.2). Assume price p_t is offered at time t to customers with patience level δ. If the customer's valuation of the product is zero at time t, then the customer leaves the system which occurs with probability $1 - d_t$. On the other hand, if the customer's valuation of the product at time t is one which occurs with probability d_t, the customer either purchases the product or waits for the next period with the following probabilities:

$$p_r(\text{purchase at time t}) = d_t \; p_r(p_t^{est} \geq 1 - \frac{1 - p_t}{\delta})$$
$$p_r(\text{wait at time t}) = d_t \; \big[1 - p_r(p_t^{est} \geq 1 - \frac{1 - p_t}{\delta})\big].$$

The $p_r(\textit{wait at time } t)$ presents the proportion of the population at time t who decided to postpone their purchase for the next period. We further assume that β fraction of the waited population from time $t-1$ stays in the system until time t. Therefore, assuming price p_{t-1} is being offered at time $t-1$, the proportion of population from time $t-1$ being added to time t is

$$\beta \; d_{t-1} \; \big[1 - p_r(p_{t-1}^{est} \geq 1 - \frac{1 - p_{t-1}}{\delta})\big].$$

Here, the β coefficient varies in $[0, 1]$. If $\beta = 0$, then no one from the waited population in time $t-1$ stays in the system. On the other hand, if $\beta = 1$, all population who decided to wait will remain in the system. The underlying assumption for our customer return model is that we have a population arrival of size one at the beginning of each period. Furthermore, we only consider waiting time of one period. The empirical expected revenue obtained in period t is computed as

$$\text{revenue} = p_r(\text{purchase at time t}) \; p_t + \beta \; p_r(\text{wait at time t} - 1) \; d_t \; p_t.$$

The first term is the revenue collected from the arrival population at time t. Moreover, the second term is the revenue obtained from the fraction of the population that waited from time $t-1$.

1.4 Strategic Thompson Sampling: Bernoulli Demand

In this section, we first present the traditional Thompson Sampling algorithm introduced to solve a stochastic multi-armed bandit problem. We then propose a Strategic Thompson Sampling algorithm that takes into account the customer's strategic behavior.

1.4.1 Traditional Thompson Sampling

The Thompson Sampling algorithm (TS) was proposed by [84] to solve the stochastic multi-armed bandit problem. In a traditional problem setting, the reward under each arm follows a Bernoulli distribution with unknown expected reward. The agent's objective is to play arms sequentially in order to find the optimal arm that gives the maximum expected reward.

Traditional Thompson Sampling

Let $S_k = 0, F_k = 0$ for all arms $k = 1, ..., N$.
Repeat the following steps for all $t = 1, ..., T$.

1. Sample Reward: Draw θ_k according to the posterior distribution $Beta(S_k + 1, F_k + 1)$.
2. Arm Selection: Pull arm i where $i = \arg\max_k \theta_k$ and observe the reward.
3. Update: If reward $= 1$, $S_i = S_{i-1} + 1$; otherwise, $F_i = F_{i-1} + 1$.

The Thompson Sampling algorithm takes a Bayesian approach and applies Beta priors on the unknown expected reward of each arm. The algorithm then updates the posterior probability in every round based on the observed reward (see Appendix B.1). Thompson Sampling is a member of the probability matching family of algorithms. This is due to the fact that, in every round, it selects the arm based on its posterior probability of being the optimal arm.

1.4.2 Strategic Thompson Sampling: Bernoulli Demand

In this section, we are proposing an updated version of the traditional Thompson Sampling algorithm that is improved in that it takes into account the customer's strategic behavior. We call this algorithm Strategic Thompson Sampling and present it as follows.

Strategic Thompson Sampling: Bernoulli Demand

Let $N_{i,0} = 0$, and $R_{i,0} = 0$ for all arms $i = 1, ..., N$. Repeat the following steps for all $t = 1, ..., T$:

1. **Retailer Price Offer Decision**

   - For each arm i, sample $\theta_i(t)$ from $Beta(1+R_{i,(t-1)}, 1+N_{i,t-1}-R_{i,t-1})$ distribution.
   - The retailer offers a price $p_{i(t)}$ that maximizes his expected empirical revenue

   $$p_{i(t)} = \underset{i}{\text{argmax}}\ p_i\ \theta_i(t).$$

2. **Customer Purchase Decision**

   - The strategic customer with patience level δ makes her purchase decision based on a Bernoulli draw with the following success probability. Given price $p_{i(t)}$ is offered in this period,

   $$\Pr(\text{success}) = d_{i(t)}\ *\ p_r\Big(p^{est} \geq 1 - \frac{1-p_{i(t)}}{\delta}\Big),$$

   where $p_r(p^{est} \geq 1 - \frac{1-p_{i(t)}}{\delta}) = \sum_i \bar{w}_i$ for all i where price p_i is revealed and $p_i \geq 1 - \frac{1-p_{i(t)}}{\delta}$.

     - If the result of the Bernoulli draw is one, the customer purchases the product.
     - If the result of the Bernoulli draw is zero, the customer does not purchase the product.

3. **Update**: The Beta parameters for the retailer are updated based on the retailer price offer and the customers' purchase decisions. If price $p_{i(t)}$ has been offered at time t, then

   - $N_{i(t),t} \leftarrow N_{i(t),(t-1)} + 1,\ N_{j,t} = N_{j,(t-1)}\ ,\ \forall j \neq i$
   - If the customer purchases immediately under price $p_{i(t)}$, then $R_{i(t),t} \leftarrow R_{i(t),(t-1)} + 1$, otherwise $R_{i(t),t} \leftarrow R_{i(t),(t-1)}$.

1.5 Numerical Results: Bernoulli Demand

In this section, we present our numerical results when the demand follows a Bernoulli distribution. In subsection 1.5.1, we plot posterior densities of expected demands for both strategic and myopic customers. We then analyze the retailer's long run price offer as customers become more and more strategic. This analysis is provided in section 1.5.2. We further numerically explore and compare the retailer's expected cumulative revenue in the presence of both myopic and strategic customers. Through extensive parameter exploration, we found a parameter setting where the retailer by applying Strategic Thompson Sampling algorithm can be better off in the presence of strategic customers. These results are presented in subsection 1.5.3. More specifically, we compare expected cumulative revenue and worst-case cumulative revenue for both types of customers in subsection 1.5.3.

1.5.1 Posterior Distributions: Bernoulli Demand

The posterior distribution of expected demand in the Strategic Thompson Sampling algorithm follows a Beta distribution in the case of Bernoulli demand. Figure 1.1 presents the density of the expected demands' posterior distributions under each price $p_1 = 0.4,\ p_2 = 0.7$, and $p_3 = 0.9$ when $T = 1000$ for both strategic and myopic customers. The true expected demand for myopic customers are $d_1 = 0.8,\ d_2 = 0.6$, and $d_3 = 0.5$; for strategic customers, the true expected demand values are $\tilde{d}_1 = 0.8,\ \tilde{d}_2 = 0.6$, and $\tilde{d}_3 = 0.3$. As we observe, the

posterior density approaches the corresponding true expected demands under each price for both types of customers.

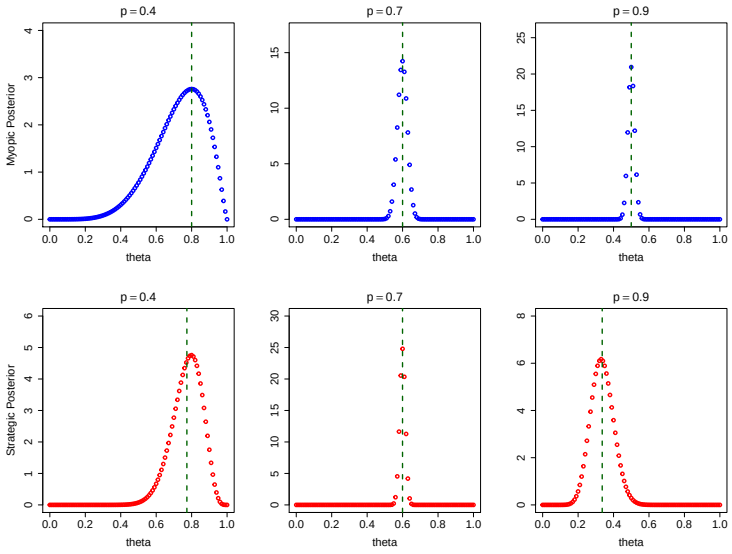

Figure 1.1: This plot represents the Beta posterior densities of the expected demands under three prices $p_1 = 0.4, p_2 = 0.7$, and $p_3 = 0.9$ when $T = 1000$. The retailer is facing myopic customers in the top-plot and strategic customers with patience level $\delta = 0.4$ in the bottom-plot. The true expected demand for myopic customers are $d_1 = 0.8,\ d_2 = 0.6$, and $d_3 = 0.5$ and for strategic customers are $\tilde{d}_1 = 0.8,\ \tilde{d}_2 = 0.6$, and $\tilde{d}_3 = 0.3$. The retailer's long-run price offers approach to $p_3 = 0.9$ in the presence of myopic customers and to $p_2 = 0.7$ for strategic customers. The green dashed-lines show the true expected demands. We averaged Beta distribution parameters over 5000 simulations.

1.5.2 Long-run Price Convergence Analysis

Our numerical results show that the convergence of the retailer price offer in the long-run depends on the value of δ, the customer's patience level. As mentioned before, the value of δ varies from 0 to 1. Customers are highly strategic for values of δ close to 1. On the other hand, for smaller values of δ, customers show less strategic and more myopic behavior. When $\delta = 0$, the customer is considered purely myopic. Figure 1.2 shows the retailer price

offer in the long-run (T=10000 periods) as δ varies from 0 to 1 in increments of 0.1. In this figure, $N = 3$, the price set is $\{p_1 = 0.4, p_2 = 0.7, p_3 = 0.9\}$, and the true expected demands for these prices are respectively, $d_1 = 0.8,\ d_2 = 0.6$, and $d_3 = 0.5$ for myopic customers.

As we observe from Figure 1.2, for highly myopic customers ($\delta \simeq 0$), the retailer's price offer converges to the highest price $p_3 = 0.9$. We observe this behavior for our parameter initialization when $0 \leq \delta \leq 0.2$. As δ increases, the retailer price offer decreases so that for more strategic customers (less myopic), the retailer-offered price converges to $p_2 = 0.7$. We observe this behavior for values of $0.3 \leq \delta \leq 0.5$. For highly strategic customers ($\delta \simeq 1$), the retailer price offer in the long-run converges to the lowest price $p_1 = 0.4$. We observe this behavior for values of $0.6 \leq \delta \leq 1$ in our set of parameter initializations. This result is consistent with the work by [15] which shows that optimal pricing strategy decreases over time in the presence of strategic customers.

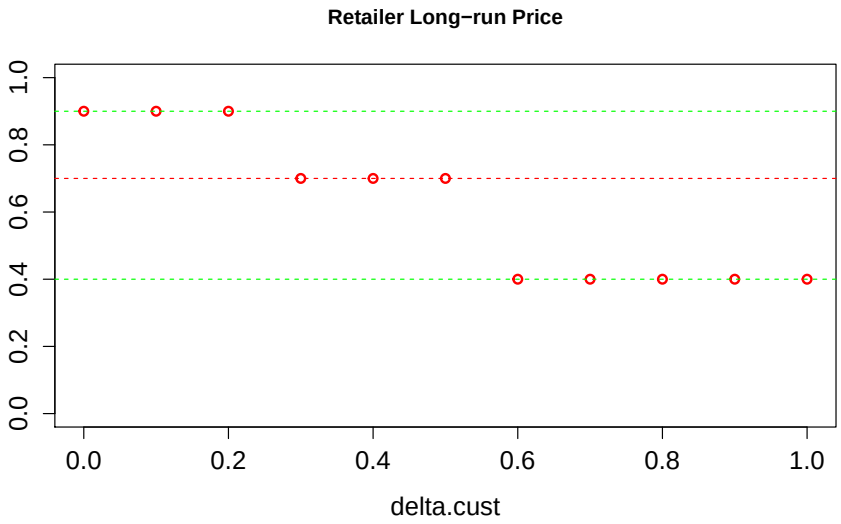

Figure 1.2: This plot presents the retailer's price offer in the long-run for $N = 3$ and T=10000. Prices for each arm are $p_1 = 0.4, p_2 = 0.7$, and $p_3 = 0.9$. The true expected demands under these prices are $d_1 = 0.8,\ d_2 = 0.6$, and $d_3 = 0.5$, respectively, in the presence of myopic customers.

1.5.3 Cumulative Revenue Analysis: Bernoulli Demand

In this section, the goal is to explore and compare the retailer's expected cumulative revenue in the presence of both myopic and strategic customers. We present a parameter initialization setting where the retailer can be better off with strategic customers if prices are offered based on the Strategic Thompson Sampling algorithm. We investigated both expected cumulative revenue and worst-case cumulative revenue scenarios. The parameter initialization for our graphs in this section are as follows. The number of arms is $N = 3$, and the length of time horizon is $T = 200$. Prices belong to the set $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$, and the expected demands in the presence of myopic customers are $d_1 = 0.9,\ d_2 = 0.5,\ d_3 = 0.3$. The strategic customers have a patience level of $\delta = 1$. The simulations are averaged over 10,000 iterations.

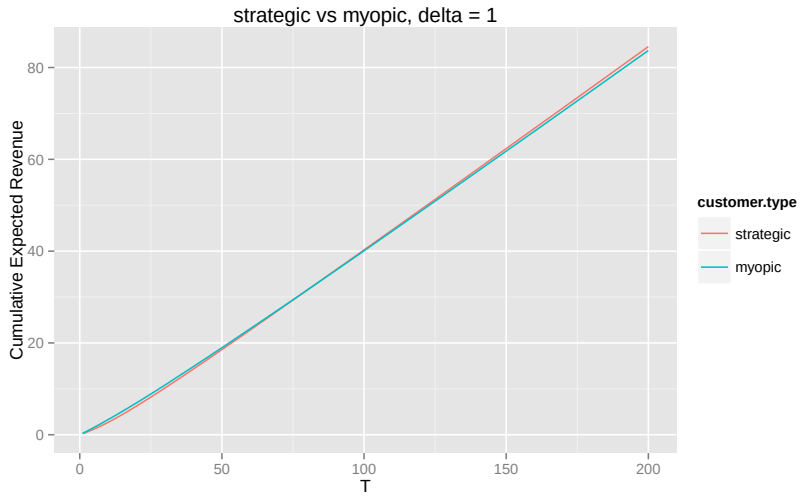

Figure 1.3: Expected cumulative revenue for both strategic and myopic customers. The retailer is better off with strategic customers after $t \geq 77$. Here, $N = 3,\ T = 200,\ \delta = 1$. The price set is $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$ and the expected cumulative demands are $d_1 = 0.9,\ d_2 = 0.5$ and $d_3 = 0.3$ in the presence of myopic customers.

Figure 1.3 shows the retailer's expected cumulative revenue. As we observe from this plot,

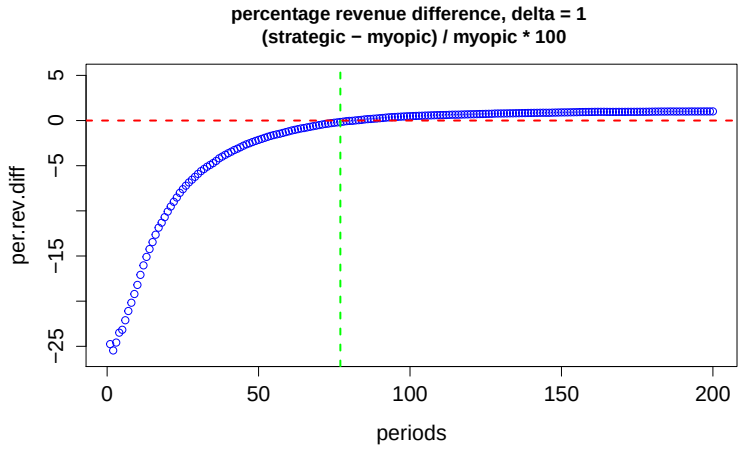

Figure 1.4: Expected cumulative revenue difference in percentages. In the long-run, the retailer's expected cumulative revenue is about 1.01% higher with strategic customers as compared to myopic ones. Here, $N = 3$, $T = 200$, $\delta = 1$. The price set is $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$ and the expected cumulative revenues are $p_1 d_1 = 0.45$, $p_2 d_2 = 0.30$, and $p_3 d_3 = 0.27$ in the presence of myopic customers.

the retailer is better off initially with myopic customers. The retailer's expected cumulative revenue is higher in the presence of strategic customers after $t \geq 77$. Figure 1.4 shows the expected cumulative revenue difference in percentages. For our set of parameter initializations, we observe that in the long run, the retailer's expected cumulative revenue is about 1.01% higher in the presence of strategic customers rather than myopic ones. Our intuition is that this behavior is due to the retailer's higher exploration of non-optimal arms in the presence of myopic customers as compared to the strategic ones. We have confirmed our intuition in Figure 1.5.

Figure 1.5 shows the average frequencies of prices offered by applying the Strategic Thompson Sampling algorithm. The average frequencies for prices $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$ are $81.1\%, 7.8\%, 11.2\%$, respectively, for myopic customers and $91.9\%, 2.6\%, 5.6\%$, respectively, for strategic customers. The long run price for both myopic and strategic customers is $p_1 = 0.5$. As we observe from this figure, the retailer chooses non-optimal arms more often

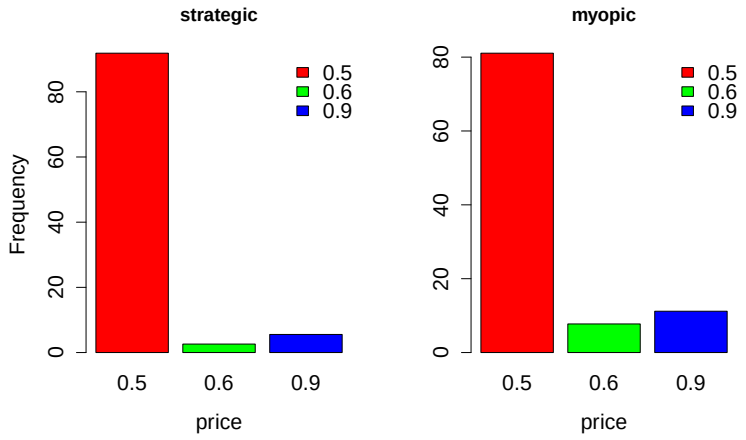

Figure 1.5: Average frequency of prices offered in T=200 periods. The average frequencies for prices $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$ are respectively $81.1\%, 7.8\%, 11.2\%$ for myopic customers and $91.9\%, 2.6\%, 5.6\%$ for strategic customers. The optimal price is $p_1 = 0.5$. The retailer chooses non-optimal arms more often in the presence of myopic customers. Simulations are averaged over 10,000 iterations.

in the presence of myopic customers as compared to the strategic ones. Here, the average frequencies are computed over 10,000 iterations.

For the rest of this section, we compute the retailer's worst-case cumulative revenue for both myopic and strategic customers for the same set of parameter initializations. Figure 1.6 shows the retailer's worst-case cumulative revenue averaged over 100,000 simulations. Figure 1.7 shows the retailer's worst-case cumulative revenue difference in percentages. As we observe from these figures, in the long run, the retailer's worst-case cumulative revenue is about 33.2% higher in the presence of strategic customers as compared to myopic ones.

1.6 Optimality of Thompson Sampling: Bernoulli Demand

Regret is defined as the expected loss due to not playing the optimal arm. In our setting, regret is the difference between optimal revenue and the revenue achieved by choosing a non-optimal price. We denote $r_i = p_i d_i$ ($\tilde{r_i} = p_i \tilde{d_i}$) as the expected revenue under arm i

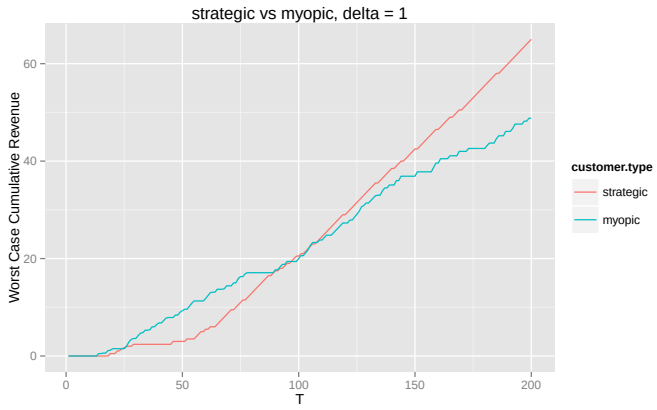

Figure 1.6: Worst-case cumulative revenue for strategic and myopic customers. The retailer is better off with strategic customers after $t > 97$. Here, $N = 3$, $T = 200$, $\delta = 1$. The price set is $p_1 = 0.5, p_2 = 0.6, p_3 = 0.9$ and the expected cumulative revenues are $p_1 d_1 = 0.45$, $p_2 d_2 = 0.30$, and $p_3 d_3 = 0.27$ in the presence of myopic customers.

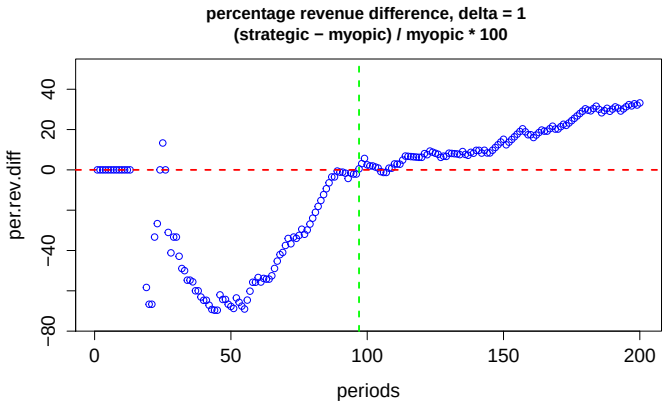

Figure 1.7: Worst-case cumulative revenue difference in percentages. In the long run, the retailer's worst case cumulative revenue is about 33.2% higher in the presence of strategic customers as compared to myopic ones.

for myopic (strategic) customers. With no loss of generality, we assume that the first arm is the optimal arm, i.e. $r_1 = r^* = \max_i \ p_i d_i$ in the presence of myopic customers and $\tilde{r}_1 = \tilde{r}^* = \max_i \ p_i \tilde{d}_i$ in the presence of strategic customers. We also assume that arm one is the unique optimal arm. Expected regret will be lower in the case of multiple optimal arms;

see [3]. We use superscript $\widetilde{()}$ to denote the strategic customer setting.

Formally, expected regret at time T is defined as

$$E[R_T^{myp}] = E\Big[\sum_{t=1}^{T}(r^* - r_{i(t)})\Big] = \sum_i \Delta_i \; E[N_{i,T}^{myp}] \tag{1.4}$$

for myopic customers and

$$E[R_T^{str}] = E\Big[\sum_{t=1}^{T}(\tilde{r}^* - \tilde{r}_{i(t)})\Big] = \sum_i \tilde{\Delta}_i \; E[N_{i,T}^{str}] \tag{1.5}$$

for strategic customers. Where, $i(t)$ denotes the arm played at time t, $\Delta_i = p_1 d_1 - p_i d_i$, and $\tilde{\Delta}_i = p_1 \tilde{d}_1 - p_i \tilde{d}_i$. Moreover, $N_{i,T}^{myp}$, and $N_{i,T}^{str}$ denote the number of plays of arm i at time T for myopic and strategic customers respectively. Equations 1.4 and 1.5 suggest that in order to bound the retailer's expected regret, we have to find an upper bound on the expected number of plays of any sub-optimal arm i, i.e., $E[N_{i,T}^{myp}]$ and $E[N_{i,T}^{str}]$. These upper bounds are stated in Theorem 1 for both types of customers. We conclude in proposition 1 that on average the retailer plays a sub-optimal arm less often in the presence of a strategic customer as compared to a myopic one. We have confirmed the result of proposition 1 empirically, as well, by looking into the average prices observed by time T; refer to Section 1.5.3, Figure 1.5.

1.6.1 Optimality of Strategic Thompson Sampling: Bernoulli Demand

In this subsection, we present our theoretical results for the Bernoulli demand scenario. More specifically, we provide upper bounds on the number of sub-optimal arm draws and compare them against both types of customers in Theorem 1 and Proposition 1, respectively. Upper bounds on the expected regret for the Strategic Thompson Sampling algorithm is presented in Proposition 2. We conclude in Proposition 3 that the retailer's expected regret is lower on average when he faces strategic customers as compared to myopic ones. We adopted the proof techniques from [3] in the Bernoulli demand scenario.

Theorem 1. [Adapted from [3], Theorem 1.] Following the Strategic Thompson Sampling algorithm, the expected number of plays of a sub-optimal arm i at time T is upper

bounded by

$$E[N_{i,T}^{myp}] = O(1) + (1+\epsilon)^2 \frac{\ln T}{KL(d_i, d_1)} + O(\frac{1}{\epsilon^2}),$$

in the presence of myopic customers and by

$$E[N_{i,T}^{str}] = O(1) + (1+\tilde{\epsilon})^2 \frac{\ln T}{KL(\tilde{d}_i, \tilde{d}_1)} + O(\frac{1}{\tilde{\epsilon}^2}),$$

in the presence of strategic customers for some $0 < \epsilon \leq 1$ and $0 < \tilde{\epsilon} \leq 1$. Here, $KL(d_i, d_1)$ is the KL-divergence between d_i, and d_1, i.e. $KL(d_i, d_1) = d_i \ln \frac{d_i}{d_1} + (1 - d_i) \ln(\frac{1-d_i}{1-d_1})$. The KL-divergence between $\tilde{d}_i$, and $\tilde{d}_1$, i.e. $KL(\tilde{d}_i, \tilde{d}_1)$ is defined in a similar fashion. Moreover, the value for ϵ is selected such that

$$(1+\epsilon)^2 \frac{\ln(T)}{KL(d_i, d_1)} = \frac{\ln(T)}{KL(x_i, y_i)},$$

where $x_i \in (d_i, d_1)$ and $y_i \in (x_i, d_1)$ are selected such that the following equations hold

$$KL(x_i, d_1) = \frac{KL(d_i, d_1)}{1+\epsilon}, \quad KL(x_i, y_i) = \frac{KL(x_i, d_1)}{1+\epsilon}.$$

The value of $\tilde{\epsilon}$ is chosen in a similar fashion.

Proof. With no loss of generality, we assume that arm one is the unique optimal arm for both strategic and myopic customers. We set x_i and y_i to be thresholds corresponding to arm i where $d_i < x_i < y_i < d_1$ holds for myopic customers and $\tilde{d}_i < x_i < y_i < \tilde{d}_1$ holds for strategic customers. We define event $E_i^{\mu}(t)$ as $\frac{R_{i,t-1}}{N_{i,t-1}+1} \leq x_i$ where $N_{i,t-1}$ and $R_{i,t-1}$ denote the number of plays of arm i and its number of successes up to time $t-1$. Event $E_i^{\theta}(t)$ is also defined as $\theta_i(t) \leq y_i$. These events show that the empirical mean and sample draw from the Beta distribution corresponding to arm i respectively are not too far from the true mean d_i ($\tilde{d}_i$). After some point in time, these events hold with high probability. Let $L_i(T) = \frac{\ln T}{d(x_i, y_i)}$. Also, let F_{t-1} denote the history of the retailer's price offers and the customer's purchase decisions up to time $t-1$. Define,

$$p_{i,t} = p_r(p_1\theta_1(t) > p_i y_i | F_{t-1})$$

and event $M_i(t)$ as

$$p_i\theta_i(t) \geq p_j\theta_j(t), \quad \forall j \neq 1$$

The expected number of plays a sub-optimal arm i at time T can be written as

$$\begin{aligned} E[N_{i,T}] &= \sum_{t=1}^{T} p_r(i(t)=i) \\ &= \sum_{t=1}^{T} p_r\Big(i(t)=i, E_i^{\mu}(t), E_i^{\theta}(t)\Big) + \sum_{t=1}^{T} p_r\Big(i(t)=i, E_i^{\mu}(t), \overline{E_i^{\theta}(t)}\Big) \\ &\quad + \sum_{t=1}^{T} p_r\Big(i(t)=i, \overline{E_i^{\mu}(t)}\Big). \end{aligned}$$

We can further provide upper bounds on each of the above terms. For comprehensive details on how to obtain these upper bounds, refer to Theorem 1 in [3].

Proposition 1. The expected number of plays of a sub-optimal arm i at time T for the Strategic Thompson Sampling algorithm is lower in the presence of a strategic customer as compared to a myopic one,

$$E[N_{i,T}^{str}] < E[N_{i,T}^{myp}].$$

Proof. We know that $KL(d_i, d_1)$ is decreasing in d_i and $\tilde{d}_i < d_i$ for any arm i. By Theorem 1, the expression for $E[N_{i,T}]$ is given as

$$E[N_{i,T}] = O(1) + (1+\epsilon)^2 \frac{\ln T}{KL(d_i, d_1)} + O(\frac{1}{\epsilon^2}).$$

Taking the derivative with respect to d_i gives

$$\begin{aligned} \frac{\partial KL(d_i, d_1)}{\partial d_i} &= \ln \frac{d_i}{d_1} + 1 - \ln \frac{1-d_i}{1-d_1} - 1 \\ &= \ln \frac{d_i(1-d_1)}{d_1(1-d_i)}. \end{aligned}$$

Since $\frac{d_i(1-d_1)}{d_1(1-d_i)} < 1$, then $\frac{\partial KL(d_i,d_1)}{\partial d_i} < 0$ and $KL(d_i, d_1)$ is decreasing in d_i. On the other hand, for the offered price $p_{i(t)}$ at time t, we have $\tilde{d}_{i(t)} = d_{i(t)} * p_r(p^{est} \geq 1 - \frac{1-p_{i(t)}}{\delta})$, i.e., $\tilde{d}_{i(t)} < d_{i(t)}$. This concludes the result in proposition 1.

Proposition 2. The Strategic Thompson algorithm provides an expected regret bound

$$E[R_T^{myp}] \leq (1+\epsilon)\sum_{i=2}^{N} \frac{\ln T}{KL(d_i, d_1)}\Delta_i + O(\frac{N}{\epsilon^2})$$

in the presence of myopic customers and

$$E[R_T^{str}] \leq (1+\epsilon)\sum_{i=2}^{N} \frac{\ln T}{KL(\tilde{d}_i, \tilde{d}_1)}\tilde{\Delta}_i + O(\frac{N}{\epsilon^2})$$

in the presence of strategic customers at time T and for the N-armed stochastic bandit problem.

Proof. The proof of this proposition is easily concluded from Theorem 1.

Proposition 3. The retailer's expected regret is lower in the presence of strategic customers as compared to myopic ones when prices are offered according to the Strategic Thompson Sampling algorithm,

$$E[R_T^{str}] < E[R_T^{str}].$$

Proof. Define $f(d_i) = \frac{\Delta_i}{KL(d_i,d_1)}$. Taking the derivative of $f(d_i)$ with respect to d_i gives

$$\frac{\partial f(d_i)}{d_i} = \ln\Big[\Big(\frac{d_1(1-d_i)}{d_i(1-d_1)}\Big)^{d_i} * \frac{1-d_1}{1-d_i}\Big] + \ln\Big[\Big(\frac{d_i(1-d_1)}{d_1(1-d_i)}\Big)^{d_i-d_1}\Big]$$

After a few steps of algebraic simplification, we get

$$\frac{\partial f(d_i)}{d_i} = \ln\Big[\Big(\frac{d_1(1-d_i)}{d_i(1-d_1)}\Big)^{d_1} * \frac{1-d_1}{1-d_i}\Big]$$

The argument of the natural logarithm is greater than one due to the fact that $KL(d_1, d_i) > 0$. Therefore, $f(d_i)$ is increasing in d_i. This combined with the regret expression given in Proposition 2 concludes the result in proposition 3.

Chapter 2

DYNAMIC PRICING FOR STRATEGIC CUSTOMERS: NORMAL DEMAND

In this chapter, we extend our model by assuming that demand follows a Normal distribution rather than the Bernoulli distribution that we analyzed in the previous chapter. We further propose and present a variation of the Strategic Thompson Sampling algorithm for the Normal demand scenario.

Similar to our Bernoulli demand model, price p_i is assigned to arm i where the expected demand and therefore expected revenue is unknown for each arm $i \in \{1, 2, ...N\}$. In order to learn the expected revenue, we apply a Bayesian framework with Gaussian prior and Gaussian likelihood as in the work by [3] with the difference that we incorporate the customer's strategic behavior and purchase decision into the model. The posterior distribution on the expected revenue for each arm is Gaussian assuming Gaussian prior and likelihood; refer to Appendix B.2. At each time t, the retailer maximizes his expected empirical revenue based on sample draws from its Gaussian posterior distribution. Then, a strategic customer with patience level δ observes the offered price and makes her purchase decision based on her utility maximization described in detail in Subsection 2.0.2. Finally, the retailer updates parameters of the posterior distribution for the selected arm.

2.0.2 Customer Purchase Decision: Normal Demand

Similar to the Bernoulli case, mentioned in Section 1.3.1, customers maximize their utility in order to make their purchase decision at time t as

$$\max \{0, v_{i(t)} - p_{i(t)}, \delta(v_{i(t)} - p^{est})\}. \tag{2.1}$$

Where $v_{i(t)}$ denotes a customer's valuation of the product under price $p_{i(t)}$. We assume that $v_i \sim uniform[0,1]$ for all prices p_i. δ measures customer's patience level, and p^{est} is the customer's estimation of the next period price. Therefore, the probability of purchase for a strategic customer with patience level δ is:

$$\begin{aligned} p_r(purchase) &= p_r(v_{i(t)} - p_{i(t)} > 0 \ , \ v_{i(t)} - p_{i(t)} > \delta(v_{i(t)} - p^{est})) \\ &= p_r(v_{i(t)} > p_{i(t)} \ , \ v_{i(t)} > \frac{p_{i(t)} - \delta\bar{p}_t}{1-\delta}) \\ &= p_r(v_{i(t)} > \max\{ \ p_{i(t)} \ , \ \frac{p_{i(t)} - \delta\bar{p}_t}{1-\delta} \ \}) \\ &= 1 - \max\{ \ p_{i(t)} \ , \ \frac{p_{i(t)} - \delta\bar{p}_t}{1-\delta} \ \}. \end{aligned} \tag{2.2}$$

We assume that customers estimate the next period price, p^{est}, as the current weighted average of observed prices, $\bar{p}_t$. We obtain the revenue at time t, i.e. r_t, as a random draw from a Normal distribution with mean $[p_r(purchase) * p_{i(t)}]$ and variance $\sigma^2_{i(t)}$. Where $p_r(purchase)$ is given in equation (2.2) and $\sigma^2_{i(t)}$ is the true variance of revenue distribution under arm $i(t)$ selected at time t. Furthermore, we know that for a myopic customer with patience level $\delta = 0$, equation (2.2) turns into $d_{i(t)} = p(v_{i(t)} > p_{i(t)})$. Since customer's valuation follows a uniform distribution on $[0,1]$ for all arms and both types of customers, $d_i = 1 - p_i$ should hold for all arms $i \in \{1, ..., N\}$.

2.0.3 Strategic Thompson Sampling: Normal Demand

In the following, we present a Strategic Thompson Sampling algorithm for the Normal demand scenario.

Let $N_{i,1} = 0$ and $\tilde{\mu}_{i,1} = 0$ for all arms $i = 1, ..., N$ at time $t = 1$.
Repeat the following steps for all $t = 1, ..., T$:

1. **Retailer Price Offer Decision**

   - Sample Revenue: For each arm i, sample $\theta_i(t)$ from the $\mathcal{N}(\tilde{\mu}_i(t), \frac{1}{N_{i,t}+1})$ distribution.

   - The retailer offers a price $p_{i(t)}$ that maximizes his expected empirical revenue,

$$p_{i(t)} = \underset{i}{\text{argmax}}\ \theta_i(t).$$

2. **Customer Purchase Decision**

   - The purchase probability for a strategic customer with patience level δ is

$$\text{p}_\text{r}(\text{purchase}) = 1 - \max\{\ p_{i(t)}\ ,\ \frac{p_{i(t)} - \delta\bar{p}_t}{1-\delta}\ \},$$

     where $\bar{p}_t$ is the average price observed by time t.

   - Revenue r_t is drawn from $N(p_r(purchase) * p_{i(t)}, \sigma^2_{i(t)})$ where $\sigma^2_{i(t)}$ is the true variance for revenue under arm $i(t)$.

3. **Update**: The retailer observes revenue r_t and updates the posterior parameters of arm $i(t)$ as

$$\tilde{\mu}_{i(t),(t+1)} = \frac{\tilde{\mu}_{i(t),t}\ N_{i(t),t} + r_t}{N_{i(t),t} + 1}, \quad N_{i(t),(t+1)} = N_{i(t),t} + 1.$$

2.1 Numerical Results: Normal Demand

In this subsection, we present numerical results for the Normal demand scenario. In our simulations, we set $N = 3$ and $T = 4000$. The price set is $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$ and expected demands are $d_1 = 0.5,\ d_2 = 0.2$, and $d_3 = 0.1$, respectively, in the presence of myopic customers. Figure 2.1 presents the Normal posterior distribution. Figure 2.2 shows the retailer's expected cumulative revenue and Figure 2.3 shows the retailer's revenue for

20 periods. Figure 2.4 shows the expected cumulative revenue difference, both in absolute values and in percentages. As we observe from these plots, the retailer's expected cumulative revenue is higher in the presence of strategic customers ($\delta = 0.99$) after $t \geq 831$. For our set of parameter initializations, we observe that in the long run, the retailer's expected cumulative revenue is about 1.8% higher in the presence of strategic customers rather than myopic ones. Our intuition is that this behavior is due to the retailer's higher exploration of non-optimal arms in the presence of myopic customers as compared to the strategic customers. We have confirmed our intuition in Figure 2.5.

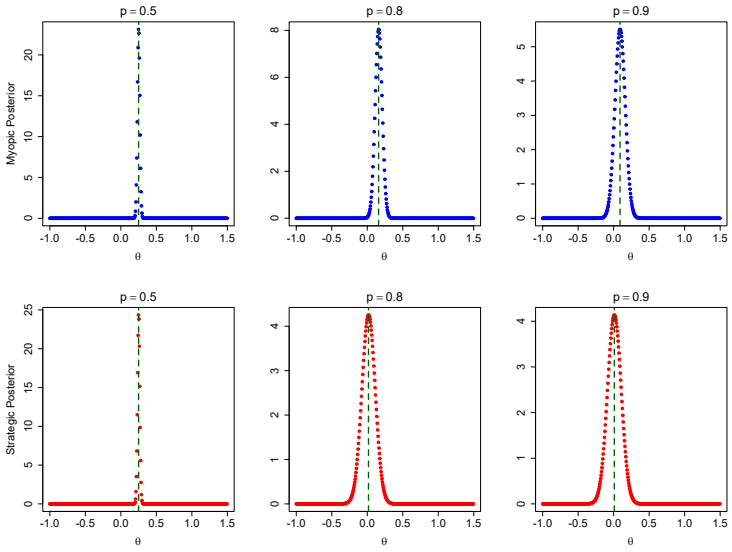

Figure 2.1: This plot represents the Gaussian posterior densities of the expected revenues at three price levels $p_1 = 0.5, p_2 = 0.8$, and $p_3 = 0.9$, when $T = 4000$. In the top plot, the retailer is facing myopic customers; in the bottom plot, the retailer faces strategic customers with patience level $\delta = 0.99$. The true expected revenues for myopic customers at the different price levels $d_1 p_1 = 0.25$, $p_2 d_2 = 0.16$, and $p_3 d_3 = 0.09$. The retailer's long run price offers approaches $p_1 = 0.5$ in the presence of both myopic and strategic customers. The green dashed lines show the true expected revenues. We averaged Normal distribution parameters over 1000 simulations.

The average frequencies of prices offered by applying a Strategic Thompson Sampling algorithm is shown in Figure 2.5. The average frequencies for prices $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$

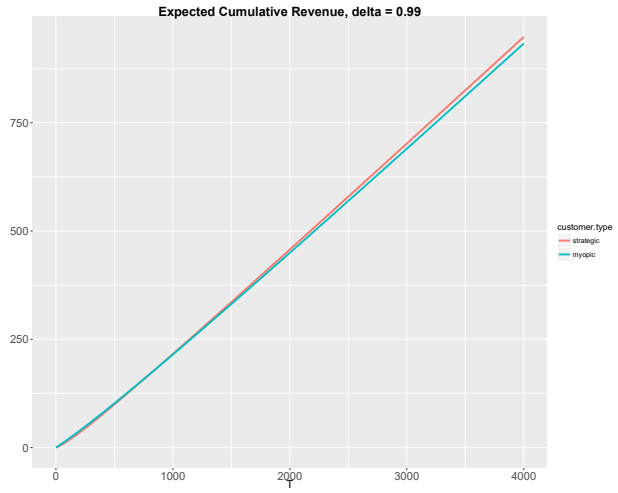

Figure 2.2: Expected cumulative revenue for both strategic and myopic customers. The retailer is better off with strategic customers after $t \geq 831$. Here, $N = 3,\ T = 4000,\ \delta = 0.99$. The price set is $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$ and the expected demands are $d_1 = 0.5,\ d_2 = 0.2$ and $d_3 = 0.1$ in the presence of myopic customers.

Figure 2.3: Revenue at each period shown only for 20 periods for both strategic and myopic customers. As observed in this plot, the retailer is better off with strategic customers around $t \geq 430$. Revenue at each time is averaged over 1000 iterations. Here, $N = 3,\ T = 4000,\ \delta = 0.99$. The price set is $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$ and the expected demands are $d_1 = 0.5,\ d_2 = 0.2$ and $d_3 = 0.1$ in the presence of myopic customers.

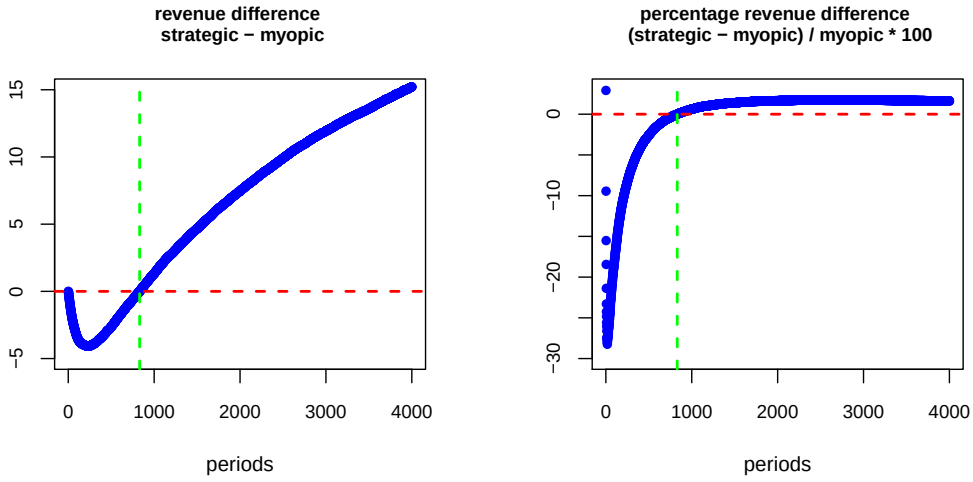

Figure 2.4: Expected cumulative revenue difference in absolute values (strategic-myopic) and in percentages are shown on the left and right plots respectively. The retailer is better off with strategic customers for $t \geq 831$. In the long run, the retailer's expected cumulative revenue is about 1.8% higher with strategic customers as compared to myopic ones. Here, $N = 3,\ T = 4000,\ \delta = 0.99$. The price set is $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$ and the expected cumulative revenues are $p_1 d_1 = 0.25,\ p_2 d_2 = 0.16$, and $p_3 d_3 = 0.09$ in the presence of myopic customers.

are respectively $85.1\%, 10.2\%, 4.8\%$ for myopic customers and $94.5\%, 2.8\%, 2.7\%$ for strategic customers. The long run price for both myopic and strategic customers is $p_1 = 0.5$. As we observe from this figure, the retailer chooses non-optimal arms more often in the presence of myopic customers as compared to the strategic ones. Here, the average frequencies are computed over 1000 iterations.

We also performed worst-case cumulative revenue analysis for both strategic and myopic customers. The numerical results are very similar to the expected cumulative revenue simulation results. Therefore, they are not reported here. We learned that the retailer is better off with strategic customers after $t > 1015$. Moreover, in the long run, the retailer's worst case cumulative revenue is about 2.6% higher in the presence of strategic customers as compared to myopic ones.

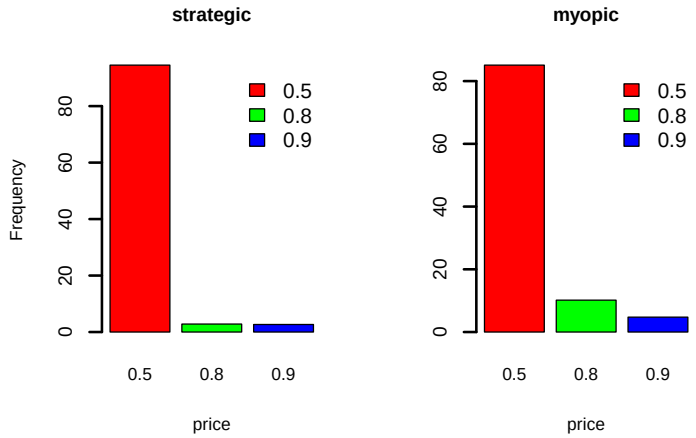

Figure 2.5: Average frequency of prices offered in T=4000 periods. The average frequencies for prices $p_1 = 0.5, p_2 = 0.8, p_3 = 0.9$ are respectively $85.1\%, 10.2\%, 4.8\%$ for myopic customers and $94.5\%, 2.8\%, 2.7\%$ for strategic customers with patience level $\delta = 0.99$. The optimal price is $p_1 = 0.5$ for both strategic and myopic customers. The retailer chooses non-optimal arms more often in the presence of myopic customers. Simulations are averaged over 1000 iterations.

2.2 Optimality of Strategic Thompson Sampling: Normal Demand

In this section, we provide analytical results for the Normal demand scenario. Specifically, in Theorem 1, we provide an upper bound on the expected number of times a suboptimal arm is drawn. Our upper bound is within a constant factor of the optimal rate based on the result provided by [51].

Theorem 1. The expected number of sub-optimal arm draws is upper bounded by

$$E[N_{a,T}] \leq 32\Delta_a^{-2} \ln T + O(1). \tag{2.3}$$

where $N_{a,T}$ denotes the number of draws of arm a by horizon time T, and $\Delta_a = \mu_1 - \mu_a$.

Proof of Theorem 1: We use the following notation in our proof: $\tilde{\mu}_{i,t} = \frac{\sum_{w=1:i(w)=i}^{t-1} r_{i(t)}}{N_{i,t}+1}$ where $\tilde{\mu}_{i,t}$ denotes the posterior mean for arm i at time t, and $r_{i(t)}$ denotes the revenue observed under offered price, $p_{i(t)}$. Moreover, $N_{i,t}$ denotes the number of times arm i is played up to

time $t-1$. By assuming a Gaussian prior distribution on the expected rewards of arm i at time t as $N(\tilde{\mu}_{i,t}, \frac{1}{1+N_{i,t}})$ and a true Gaussian likelihood of $N(\mu_i, 1)$, we derive the posterior distribution as $N(\tilde{\mu}_{i,t+1}, \frac{1}{1+N_{i,t+1}})$. At each time step, the Strategic Thompson Sampling algorithm with Gaussian prior draws independently from the posterior distributions of each arm, and then pulls the arm with the maximum random draw. In our proof, $\theta_{i,t}$ denotes the random draw from the Gaussian posterior distribution of arm i at time t. In this subsection, we determine an upper bound on the expected number of plays of a sub-optimal arm and the expected regret of strategic Thompson Sampling algorithm with Gaussian priors.

Let $E[N_{a,T}]$ denote the average number of plays of a sub-optimal arm a up to horizon time T. In the following, we present the derivation of an upper bound on $E[N_{a,T}]$. The sample draw from the optimal arm is either underestimated or concentrated around its true mean, whenever a sub-optimal arm is being selected. We bound $E[N_{a,T}]$ by obtaining upper bounds on these events.

$$E[N_{a,T}] = \sum_{t=1}^{T} p_r[\theta_1(t) \leq \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a] + \sum_{t=1}^{T} p_r[\theta_1(t) > \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a] \tag{2.4}$$

where β is constant and its value will be determined throughout the proof. In the following two subsections, we find upper bounds for the first term and second term in equation (2.4) respectively. Throughout the proof, we apply the following tail bound for a Gaussian random variable $Z \sim N(\mu, \sigma^2)$, $p_r(\ Z - \mu \leq -\sqrt{2\sigma^2 x}\ ) \leq e^{-x}$.

2.2.1 Upper bound on the first term

In this subsection, we find an upper bound on the first term in equation (2.4) where $\theta_{1,t}$ follows a Normal distribution, $\theta_{1,t} \sim N(\tilde{\mu}_{1,t}, \tilde{\sigma}^2_{1,t})$ with $\tilde{\sigma}^2_{1,t} = \frac{1}{1+N_{1,t}}$.

$$p_r\Big[\theta_{1,t} \leq \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a\Big] = p_r\Big[(\theta_{1,t} - \tilde{\mu}_{1,t}) \leq (\mu_1 - \tilde{\mu}_{1,t}) - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a\Big]$$

Define random variable $U := \mu_1 - \tilde{\mu}_{1,t}$, and constant $b := \sqrt{\frac{\beta \ln t}{1+N_{1,T}}}$. The above probability can be written as

$$\begin{aligned}
p_r(\theta_{1,t} - \tilde{\mu}_{1,t} \leq U - b) &= E_{U\sim u}\left[p_r\Big(\theta_{1,t} - \tilde{\mu}_{1,t} \leq -(b-u)|U=u\Big)\right] \\
&\leq E_{U\sim u}\left[e^{-\frac{1+N_{1,t}}{2}(b-u)^2}\mathbb{I}\Big\{b-u \geq \sqrt{\frac{2\ln t^\gamma}{1+N_{1,t}}}\Big\} + \mathbb{I}\Big\{b-u < \sqrt{\frac{2\ln t^\gamma}{1+N_{1,t}}}\Big\}\right] \\
&\leq t^{-\gamma} + p_r\Big[\sqrt{\frac{\beta\ln t}{1+N_{1,t}}} - (\mu_1 - \tilde{\mu}_{1,t}) < \sqrt{\frac{2\ln t^\gamma}{1+N_{1,t}}}\Big] \\
&\leq t^{-\gamma} + p_r\Big[\mu_1 - \tilde{\mu}_{1,t} > \sqrt{\frac{\beta\ln t}{1+N_{1,t}}} - \sqrt{\frac{2\ln t^\gamma}{1+N_{1,t}}}\Big] \\
&\leq t^{-\gamma} + p_r\Big[\mu_1 - \tilde{\mu}_{1,t} > \frac{1}{2}\frac{\frac{\beta\ln t}{1+N_{1,t}} - \frac{2\ln t^\gamma}{1+N_{1,t}}}{\sqrt{\frac{2\ln t^\gamma}{1+N_{1,t}}}}\Big] \qquad (2.5)
\end{aligned}$$

The last inequality follows from Taylor series expansion of $\sqrt{x}$ around a,

$$\sqrt{x} - \sqrt{a} > \frac{1}{2}\frac{x-a}{\sqrt{a}}, \quad \forall x > a.$$

After some algebraic simplification, (2.5) becomes

$$\begin{aligned}
p_r(\theta_{1,t} - \tilde{\mu}_{1,t} \leq U - b) &\leq t^{-\gamma} + p_r(\mu_1 - \tilde{\mu}_{1,t} > \frac{1}{2}\frac{1}{\sqrt{1+N_{1,t}}}\ln t^{\beta-2\gamma}(\ln t^{2\gamma})^{-\frac{1}{2}}) \\
&= t^{-\gamma} + p_r(\tilde{\mu}_{1,t} - \mu_1 < -\sqrt{2\tilde{\sigma}^2_{1,t}X})
\end{aligned}$$

where $\sqrt{X} := \frac{1}{2\sqrt{2}}\ln t^{\beta-2\gamma}(\ln t^{2\gamma})^{\frac{-1}{2}}$, $X = \frac{(\beta-2\gamma)^2}{16\gamma}\ln t$, and $\tilde{\sigma}^2_{1,t} = \frac{1}{1+N_{1,t}}$. By applying Gaussian tail bound, we obtain

$$p_r\Big[\theta_{1,t} \leq \mu_1 - \sqrt{\frac{\beta\ln t}{1+N_{1,t}}}, A(t) = a\Big] \leq t^{-\gamma} + t^{-\frac{(\beta-2\gamma)^2}{16\gamma}}$$

Summing over all times t gives

$$\sum_{t=1}^{T} p_r\Big[\theta_{1,t} \leq \mu_1 - \sqrt{\frac{\beta\ln t}{1+N_{1,t}}}, A(t) = a\Big] \leq \sum_{t=1}^{T} t^{-\gamma} + \sum_{t=1}^{T} t^{-\frac{(\beta-2\gamma)^2}{16\gamma}} \qquad (2.6)$$

For the two sum terms in equation (2.6) to be convergent and of order of one, we need:

$$\gamma > 1,\ \frac{(\beta-2\gamma)^2}{16\gamma} > 1 \quad \Rightarrow \quad \gamma > 1,\ \beta > 2\gamma + 4 \quad \Rightarrow \quad \gamma > 1,\ \beta > 6$$

Set $\gamma = 1+\epsilon$, and $\beta = 6(1+\epsilon)$ where $\epsilon > 0$ is an arbitrary small number,

$$\sum_{t=1}^{T} p_r\Big[\theta_1(t) \leq \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a\Big] \leq 2\sum_{t=1}^{T} t^{-(1+\epsilon)} = O(1).$$

To conclude, the first term in equation (2.4) is of constant order,

$$\sum_{t=1}^{T} p_r\Big[\theta_1(t) \leq \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a\Big] = O(1). \tag{2.7}$$

2.2.2 Upper bound on the second term

The next step is to provide an upper bound for the second term in equation (2.4). We first condition on the event that the optimal arm is selected enough number of times. Specifically, whether $N_{1,t} > \alpha\Delta_a^{-2}\ln t - 1$ where $\Delta_a = \mu_1 - \mu_a$ and α is a constant number which will be determined towards the end of the proof.

$$\begin{aligned}
&\sum_{t=1}^{T} p_r\Big[\theta_{1,t} > \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a\Big] \\
&\qquad \leq \sum_{t=1}^{T} p_r\Big[\theta_{a,t} > \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t) = a, N_{1,t} \geq \alpha\Delta_a^{-2}\ln t - 1\Big] \\
&\qquad\quad + \sum_{t=1}^{T} p_r\Big[N_{1,t} < \alpha\Delta_a^{-2}\ln t - 1\Big] \\
&\qquad \leq \sum_{t=1}^{T} p_r\Big[\theta_{a,t} > \mu_1 - \sqrt{\frac{\beta}{\alpha}}\,\Delta_a, A(t) = a\Big] + \sum_{t=1}^{T} E\Big[\mathbb{I}\{N_{1,t} < \alpha\Delta_a^{-2}\ln t - 1\}\Big] \\
&\qquad \leq \sum_{t=1}^{T} p_r\Big[\theta_{a,t} > \mu_1 - \sqrt{\frac{\beta}{\alpha}}\,\Delta_a, A(t) = a\Big] + E\Big[\sum_{t=1}^{T}\mathbb{I}\{N_{1,t} < \alpha\Delta_a^{-2}\ln T\}\Big] \\
&\qquad \leq \sum_{t=1}^{T} p_r\Big[\theta_{a,t} > \mu_1 - \sqrt{\frac{\beta}{\alpha}}\,\Delta_a, A(t) = a\Big] + \alpha\Delta_a^{-2}\ln T
\end{aligned}$$

In the second inequality, we apply the fact that $\theta_{a,t} > \theta_{1,t}$ since arm a is being pulled at time t. Here, we set parameter α, and β such that

$$\mu_1 - \sqrt{\frac{\beta}{\alpha}}\,\Delta_a > \frac{\mu_1+\mu_a}{2} \quad\Rightarrow\quad \frac{1}{2}\Delta_a > \sqrt{\frac{\beta}{\alpha}}\Delta_a \quad\Rightarrow\quad \alpha > 4\beta.$$

Therefore,

$$\begin{aligned}
&\sum_{t=1}^{T} p_r\Big[\theta_{1,t} > \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t)=a\Big] \\
&\qquad \le \sum_{t=1}^{T} p_r\Big[\theta_{a,t} > \frac{\mu_1+\mu_a}{2}, A(t)=a\Big] + \alpha\Delta_a^{-2}\ln T \\
&\qquad \le \sum_{t=1}^{T} p_r\Big[\theta_{a,t} - \mu_a > \frac{\Delta_a}{2}, A(t)=a\Big] + \alpha\Delta_a^{-2}\ln T \\
&\qquad \le \sum_{t=1}^{T} E\Big[e^{-\frac{1}{2}\frac{\Delta_a^2}{4}\frac{N_{a,t}+1}{2}}\mathbb{I}\Big\{N_{a,t} \ge 16\Delta_a^{-2}\ln T\Big\}\Big] \\
&\qquad\quad + \sum_{t=1}^{T} p_r\Big(N_{a,t} < 16\Delta_a^{-2}\ln T, \theta_{a,t} > \frac{\mu_1+\mu_a}{2}\Big) + \alpha\Delta_a^{-2}\ln T \\
&\qquad \le 1 + \sum_{t=1}^{T} p_r\Big(\theta_{a,t} > \frac{\mu_1+\mu_a}{2} | N_{a,t} < 16\Delta_a^{-2}\ln T\Big) \\
&\qquad\quad \times\ E\Big[\mathbb{I}\Big\{N_{a,t} < 16\Delta_a^{-2}\ln T\Big\}\Big] + \alpha\Delta_a^{-2}\ln T \\
&\qquad \le 1 + \frac{1}{2}E\Big[\sum_{t=1}^{T}\mathbb{I}\Big\{N_{a,t} < 16\Delta_a^{-2}\ln T\Big\}\Big] + \alpha\Delta_a^{-2}\ln T \\
&\qquad \le 1 + 8\Delta_a^{-2}\ln T + \alpha\Delta_a^{-2}\ln T.
\end{aligned}$$

To summarize, for $\alpha > 4\beta$, the second term in equation (2.4) is upper bounded by

$$\sum_{t=1}^{T} p_r\Big[\theta_{1,t} > \mu_1 - \sqrt{\frac{\beta \ln t}{1+N_{1,t}}}, A(t)=a\Big] \le 1 + (8+\alpha)\Delta_a^{-2}\ln T. \tag{2.8}$$

To conclude, by setting $\alpha = 24$, and applying the upper bounds for the first and second terms in equations (2.7), and (2.8) respectively, we obtain

$$E[N_{a,T}] \le 32\Delta_a^{-2}\ln T + O(1). \tag{2.9}$$

This completes our proof of theorem 1.

2.2.3 Lower bound on the second term

Our objective for this subsection is to investigate the lower bound on the second term in equation (2.4) by assuming that arm one is selected an infinite number of times. We further prove asymptotic optimal rate for the lower bound term.

$$\sum_{t=1}^{T} p_r[\theta_1(t) > \mu_1, A(t) = a] < \sum_{t=1}^{T} p_r[\theta_1(t) > \mu_1 - \sqrt{\frac{6 \ln t}{1 + N_{1,t}}}, A(t) = a]. \tag{2.10}$$

In the following, we derive an upper bound on the left-hand side term in equation (2.10).

$$\sum_{t=1}^{T} p_r(\theta_{1,t} > \mu_1, A(t) = a) \leq \sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1, N_{a,T} \geq s_T) + \sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1, N_{a,T} < s_T) \tag{2.11}$$

The first term in (2.11) is upper bounded by:

$$\begin{aligned}
\sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1, N_{a,T} \geq s_T) &\leq \sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1 | N_{a,T} \geq s_T)\ p_r(N_{a,T} \geq s_T) \\
&\leq \sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_a + \mu_1 - \mu_a | N_{a,T} \geq s_T) \\
&\leq \sum_{t=1}^{T} e^{-\frac{(\mu_1 - \mu_a)^2}{4} s_T}
\end{aligned}$$

By setting $s_T := 4(\mu_1 - \mu_a)^{-2} \log(T)$, we get

$$\sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1, N_{a,T} \geq s_T) \leq 1.$$

The second term in (2.11) is upper bounded by

$$\begin{aligned}
\sum_{t=1}^{T} p_r(\theta_{a,t} > \mu_1, N_{a,T} < s_T) &= \sum_{t=1}^{T} p_r(N_{a,T} < s_T | \theta_{a,t} > \mu_1)\ p_r(\theta_{a,t} > \mu_1) \\
&= \sum_{t=1}^{T} E\Big[I\{N_{a,T} < s_T | \theta_{a,t} > \mu_1\} \Big]\ p_r(\theta_{a,t} > \mu_1) \\
&\leq \frac{1}{2} E\Big[\sum_{t=1}^{T} I\{N_{a,T} < s_T | \theta_{a,t} > \mu_1\} \Big] \\
&\leq \frac{1}{2} s_T.
\end{aligned}$$

In the second to last inequality, we use $p_r(\theta_{a,t} > \mu_1) < \frac{1}{2}$. This is due to the fact that $\theta_{a,t}$ follows a Gaussian distribution with mean and median equal to $\mu_{a,t}$ which are less than μ_1. Therefore,

$$\sum_{t=1}^{T} p_r(\theta_{1,t} > \mu_1, A(t) = a) \leq 1 + 2(\mu_1 - \mu_a)^{-2} \ln(T) = 1 + \frac{\ln(T)}{KL(\mu_1, \mu_a)}. \tag{2.12}$$

The bound given above is asymptotically optimal based on the result in [51].

2.3 Retailer's Performance Comparison: Myopic vs Strategic

Following the Strategic Thompson Sampling algorithm and applying the results in Theorem 1, the expected number of plays of a sub-optimal arm i at time T is upper bounded by

$$E[N_{i,T}^{myp}] \leq \frac{32 \ln T}{(\mu_1 - \mu_i)^2} + O(1), \tag{2.13}$$

in the presence of myopic customers and by

$$E[N_{i,T}^{str}] \leq \frac{32 \ln T}{(\tilde{\mu}_1 - \tilde{\mu}_i)^2} + O(1), \tag{2.14}$$

in the presence of strategic customers. Here, $\frac{1}{KL(\mu_i,\mu_1)} = \frac{2}{(\mu_1-\mu_i)^2}$, and $\frac{1}{KL(\tilde{\mu}_i,\tilde{\mu}_1)} = \frac{2}{(\tilde{\mu}_1-\tilde{\mu}_i)^2}$ due to Pinsker's inequality for the case of equal variances (see Appendix A).

Proposition 1. For the case of equal optimal arms for both types of customers, the expected number of plays of a sub-optimal arm i at time T for the Strategic Thompson Sampling algorithm is lower in the presence of strategic customers as compared to myopic ones,

$$E[N_{i,T}^{str}] < E[N_{i,T}^{myp}].$$

Proof. For the offered price p_i at time t, we have

$$\begin{aligned}
\tilde{d}_i &= 1 - \max\{p_i, \frac{p_i - \delta \bar{p}_t}{1 - \delta}\} \\
&= d_i \, \mathbb{I}\{p_i < \bar{p}_t\} + (d_i - \frac{\delta(p_i - \bar{p}_t)}{1 - \delta}) \, \mathbb{I}\{p_i > \bar{p}_t\} \\
&\leq d_i
\end{aligned}$$

Therefore, $\tilde{\mu}_i = p_i\tilde{d}_i < \mu_i = p_i d_i$ for all the suboptimal arms. Therefore, for the case of $\mu_1 = \tilde{\mu}_1$, $\frac{1}{(\tilde{\mu}_1 - \tilde{\mu}_i)^2} < \frac{1}{(\mu_1 - \mu_i)^2}$. This along with the asymptotically optimal expressions for the average number of sub-optimal arms' draws conclude the result in this proposition.

Proposition 2. The Strategic Thompson algorithm provides an expected regret bound

$$E[R_T^{myp}] \leq \sum_{i=2}^{N} \frac{32 \ln T}{(\mu_1 - \mu_i)} + O(N)$$

in the presence of myopic customers and

$$E[R_T^{str}] \leq \sum_{i=2}^{N} \frac{32 \ln T}{(\tilde{\mu}_1 - \tilde{\mu}_i)} + O(N)$$

in the presence of strategic customers at time T and for the N-arm stochastic bandit problem.

Proof. The proof of this theorem is easily adapted from theorem 1. Since the expected regret for one play of sub-optimal arm i, is $\mu_1 - \mu_i$ for the myopic customer and $\tilde{\mu}_1 - \tilde{\mu}_i$ for the strategic customer.

Proposition 3. For the case of equal optimal arms for both types of customers, the retailer's expected regret is lower in the presence of strategic customers as compared to myopic ones when prices are offered according to the Strategic Thompson Sampling algorithm,

$$E[R_T^{str}] < E[R_T^{myp}].$$

Proof. This result follows from the regret expression given in Proposition 2 and the relation between $\tilde{\mu}_i$ and μ_i where $\tilde{\mu}_i < \mu_i$ for any sub-optimal price p_i.

Chapter 3

BAYESIAN ONLINE LEARNING WITH EMPIRICAL BAYES PRIOR

3.1 Introduction

Sponsored search advertising helps providers to bring traffic from search engines to content provider websites. The majority of search engines' revenues come from this advertising business model. The overall search ad revenue in the US was $36.69 billion dollars in 2017; this number is expected to grow by 24% in 2019 according to an article in Search Engine Land [1]. Initially, all the interested advertisers introduce keywords along with their corresponding bids for them. Every time a search query is requested by the user, the search engine has to fill out a fixed number of slots available for advertisements that are relevant to the users' search query. Before rendering the search results, the search engine plays an auction referred to as a keyword auction to select the most profitable and relevant advertisements for the ad slots. Then, advertisers pay the search engine, but only if the user clicks on the ad (this payment scheme is referred to as "pay-per-click").

Therefore, obtaining an accurate click-through-rate (CTR) prediction is crucial to the search engine, since it can generate higher revenue and a better search engine user experience. [42] built a Bayesian linear probit model to predict click-through-rate. They assumed factorizing normal prior distribution on feature weights. They approximated the posterior distributions using an expectation propagation algorithm ([63]). They then provided closed-form update equations for the parameters of the posterior distributions (mean and variance of the Gaussians). The algorithm initially starts with a standard Normal distribution as the prior belief,

[1]The article can be found here: https://searchengineland.com/google-search-ad-revenues-271188

and computes updated posterior parameters for the first training sample. Thereafter, they use the computed posterior as the prior for the next training data point and iterate.

Our objective in this chapter is to build an informative prior – and more specifically, an empirical Bayes prior – for the Bayesian online learning algorithm that performs binary prediction. The underlying model used in this chapter is the same as the one in [42] with a difference that we are applying the Bayesian Linear Probit (BLIP) model to binary classification problem on a public data set called "Census Income Data Set". The output variable denotes whether a particular person earns more than $50,000 or not. Our main objective is to build an informative prior using a portion of the training data set and start the BLIP model with the built-in prior rather than the non-informative standard Normal distributions. We further compare the prediction accuracies of the BLIP model with informative and non-informative priors. An empirical Bayes model (BLIP with empirical Bayes prior) has been implemented recently in the production system of one of the largest online retailers. The web-lab experiment is currently running.

The rest of this chapter is structured as follows. In Section 3.2, we briefly describe the BLIP model. We explain Empirical Bayes methodology in details and derive the empirical Bayes prior variance in Section 3.3. Detailed description of our data set is included in Section 3.4. We report empirical Bayes variance simulations in Section 3.5. Prediction accuracies of BLIP model with and without informative priors are reported in Section 3.6. More simulations are presented in Section 3.7. We conclude this chapter with potential future directions in Section 3.8.

3.2 Model Description

3.2.1 Bayesian Linear Probit Model

In this subsection, we first provide a brief summary of Bayesian Linear Probit (BLIP) model ([42]) which is used throughout this chapter. We assume that the true probability distribution

is given as

$$p(y|x,w) = \Phi(\frac{y.w^T x}{\beta})$$

where $\Phi(.)$ is the CDF of the standard Gaussian distribution (inverse link function) and β scales its steepness. Here, $y \in \{1,-1\}$ denotes the click/non-click label, where 1 represents a click and -1 a non-click. Moreover, x denotes the vector of various types of features including ad, query, and context that are used to predict the click rate for an ad impression. In this study, we have N categorical features, where feature i can take M_i values. We represent x as a sparse binary feature vector

$$x = (x_1^T, ..., x_N^T)^T, \quad x_i^T = (x_{i,1}, ..., x_{i,M_i}), \quad \sum_{j=1}^{M_i} x_{i,j} = 1.$$

In order to learn the weights "w" in a Bayesian framework, we assume a factorizing Gaussian prior distribution as

$$p(w) = \prod_{i=1}^{N} \prod_{j=1}^{M_i} N(w_{i,j}; \mu_{i,j}, \sigma_{i,j}^2)$$

where $w_{i,j}$ denotes the weight for feature i-th with j-th value. The posterior distribution using the likelihood and prior distribution is given as

$$p(w|x,y) \propto p(y|x,w)p(w).$$

Approximation of the posterior can be obtained using approximate message passing since there is no closed form solution for it. For more details refer to [42] and the expectation propagation algorithm introduced in [63]. Assume that the vector of mean and variances are give as

$$\mu := (\mu_{1,1}, ..., \mu_{N,M_N})^T, \quad \sigma^2 := (\sigma_{1,1}^2, ..., \sigma_{N,M_N}^2)^T$$

The posterior parameter update equations are

$$\tilde{\mu}_{ij} = \mu_{ij} + y x_{ij} \frac{\sigma_{ij}^2}{\Sigma} v(\frac{y.x^T \mu}{\Sigma})$$
$$\tilde{\sigma}_{ij}^2 = \sigma_{ij}^2 [1 - x_{ij} \frac{\sigma_{ij}^2}{\Sigma} w(\frac{y.x^T \mu}{\Sigma})]$$

where $v(t) := \frac{N(t;0,1)}{\phi(t;0,1)}$ and $w(t) := v(t)(v(t)+t)$.

3.3 Empirical Bayes: BlipBayes Model

One potential approach for building an informative prior is to apply empirical Bayes methodology (see [31]). In a hierarchical Bayesian setting, applying empirical Bayes means estimating hyper-parameters using all/portion of the available data. In this subsection, we explain the steps for deriving an empirical Bayes prior variance formula. Assume we have N features of the same order (either all first order features or interaction terms). Categorical feature i can have at most M_i categories. Furthermore, w_{ij} denotes the average weight for feature ith, in jth category. As an example in Census Income Data Set, occupation is a categorical feature with 14 categories such as Sale, Prof-specialty, Exec-managerial, etc.

Our objective is to have a Bayesian framework to learn the average weight μ_{ij}. We assume that the unknown distribution on all weights in a particular order (1st or 2nd) follows a Normal distribution with unknown parameters, mean μ and variance τ^2,

$$\mu_{ij} \sim N(\mu, \tau^2) \qquad \forall i, j.$$

We set the mean μ to zero to ensure the model is invariant to contextual sign changes. The question is to come up with a formula to estimate τ^2, empirical Bayes prior variance. After time $t = 1$, we can obtain an estimate of μ_{ij}^1 as follows:

$$\mu_{ij}^1 | \mu_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$$

By having

$$\begin{aligned} E[\mu_{ij}^1|\mu_{ij}] = \mu_{ij},\ Var[\mu_{ij}^1|\mu_{ij}] = \sigma_{ij}^2,\ E[\mu_{ij}^1] = \mu \\ Var[\mu_{ij}^1] = E[Var[\mu_{ij}^1|\mu_{ij}] + Var(E(\mu_{ij}^1|\mu_{ij})) \\ = \frac{\Sigma_{ij}\sigma_{ij}^2}{N} + \tau^2 \end{aligned}$$

The second equation comes from the variance decomposition. On the other side, by running a BLIP model starting from $N(0, 1)$ prior distribution after $t = 1$, we get the posterior

distributions on μ_{ij} as $N(\hat{\mu}_{ij}, \hat{\sigma}^2_{ij})$ which in turn can give us another estimate for $Var(\mu^1_{ij})$ as

$$Var(\mu^1_{ij}) \simeq \frac{\sum_{ij}(\hat{\mu}_{ij} - \mu)^2}{N} = \frac{\sum_{ij} \hat{\mu}^2_{ij}}{N}.$$

The last equality comes from setting $\mu = 0$ as mentioned above. By comparing the last two equations we obtain

$$\tau^2 \simeq \frac{\sum_{ij}(\hat{\mu}_{ij} - \mu)^2}{N} - \frac{\sum_{ij} \hat{\sigma}^2_{ij}}{N} = \frac{\sum_{ij}[(\hat{\mu}_{ij} - \mu)^2 - \hat{\sigma}^2_{ij}]}{N} = \frac{\sum_{ij}[\hat{\mu}^2_{ij} - \hat{\sigma}^2_{ij}]}{N} \tag{3.1}$$

3.4 Data Set Description

The data set used in this study is known as "Adult Data Set" or "Census Income Data Set" which is publicly available [2]. The objective is to predict whether income exceeds $50,000 per year based on census data. The train data set has 30,162 observations, and the test data set has 15,060 observations after removing rows with empty feature values. We have, in total, 13 first order categorical features. We are also adding all the interaction terms (total of 78) to the data set in order to compute two empirical Bayes variances for first and second order features.

Figure 3.1 shows the distribution of four categorical features: work class, education, occupation, and marital status in the train data set. The following table gives summary statistics of the data.

3.5 Empirical Bayes Variance Simulations

Including all first and second order features gave negative τ^2 values for both feature orders. After training the BLIP model for the first day, we got almost all of the Gaussian posterior weights with mean and variance close to zero. This was such that the σ^2 of the Gaussian posterior was larger than its μ^2 for almost all weights. Based on the empirical Bayes variance

[2]Link to data set: "https://archive.ics.uci.edu/ml/datasets/adult".

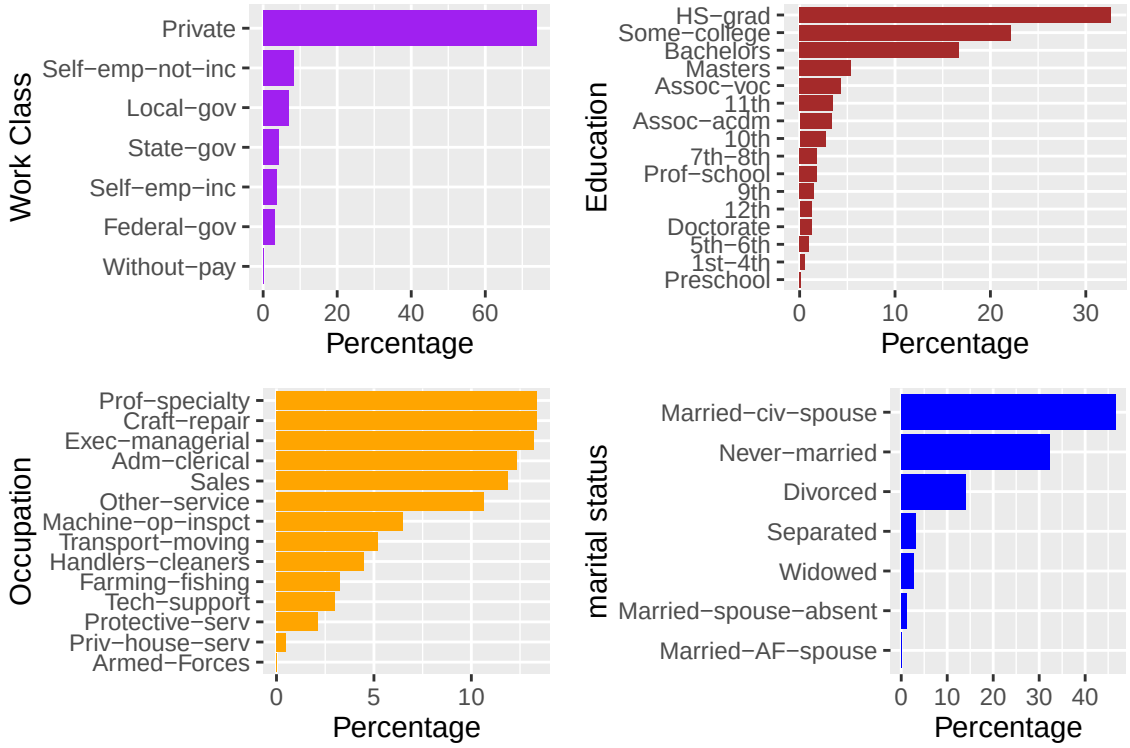

Figure 3.1: The distribution of four categorical features work class, education, occupation, and marital status in the train data set.

formula given in equation (3.1), this results in a negative value for τ^2. This section reports simulation ideas and their numerical values for empirical Bayes prior variances τ_1^2 and τ_2^2 for first and second order features respectively.

3.5.1 Feature selection

The first idea was to perform feature selection. We ran regularized linear regression with lasso, adaptive lasso, and elastic net using the "glmnet" package in R-language. We applied 10−fold cross validation to choose the shrinkage parameter. We then ran BLIP model for the first day and computed empirical Bayes variances in two scenarios, one with selected features and one with all first-order features and only the selected second-order ones. These

Feature	numCat	Feature	numCat
occupation	14	workclass	7
education	16	education-num	10
relationship	6	marital status	7
gender	2	race	5
hours-per-week	12	native-country	41
capital gain	3	capital loss	2
age	20		

Table 3.1: This table lists all first order categorical features along with their numbers of categories observed.

simulations returned a negative value for τ^2. The objective functions for our feature selection methodologies lasso, adaptive lasso, and elastic net are given as

$$\hat{W}(lasso) = \arg\min_{w} \ ||y - XW||^2 + \lambda \sum_{i,j} |w_{i,j}|$$

$$\hat{W}(adaptive\ lasso) = \arg\min_{w} \ ||y - XW||^2 + \lambda \sum_{i,j} \hat{\tau}_{i,j} |w_{i,j}|$$

$$\hat{W}(elastic\ net) = \arg\min_{w} \ ||y - XW||^2 + \lambda_1 \sum_{i,j} |w_{i,j}| + \lambda_2 \sum_{i,j} w_{i,j}^2$$

where λ is the shrinkage parameter estimated using 10-fold cross validation. $w_{i,j}$ is the weight coefficient for i-th feature in j-th category. $\hat{\tau}$ in the last equation is the adaptive weight vector. It is defined as $\hat{\tau}_{i,j} = \frac{1}{(\hat{w}_{i,j}^{ini})^{\gamma}}$ where $\hat{w}_{i,j}^{ini}$ is an initial estimate of $w_{i,j}$ coefficient. We obtained these initial estimates by performing ridge regression first. Moreover, γ is a positive constant that adjusts the adaptive lasso weight vector and takes values 0.5, 1, and 2. We set $\gamma = 1$ in our simulations. Based on $\hat{\tau}_{i,j}$ formula, coefficients with low initial estimates get penalized more by adaptive lasso.

3.5.2 Linear regression per feature

As another approach to estimate empirical Bayes variances, we ran linear regression per feature and used the coefficient estimate and the square of its standard error as estimates for the mean and variance in τ^2 formula given in (3.1). We iterated this process over all features to compute τ^2. We also explored the feature selection approach first, and then ran linear regression per each selected feature to compute τ^2. The numerical values for empirical Bayes variances τ_1^2 and τ_2^2 are summarized below. In all scenarios, we obtained small positive values for empirical Bayes variances of both first and second order features.

For a train data set of 5000 samples, we first performed feature selection by applying lasso, adaptive lasso or elastic net and then ran linear regression per each selected feature. We obtained empirical Bayes variances $\tau_1^2 = 0.03037648$ an d$\tau_2^2 = 0.02936641$ in the case of lasso, and $\tau_1^2 = 0.069$ and $\tau_2^2 = 0.029$ with adaptive lasso. For elastic net with $\alpha = 0.5$, we obtained $\tau_1^2 = 0.027$, and $\tau_2^2 = 0.029$ and with $\alpha = 0.8$, we obtained $\tau_1^2 = 0.029$ and $\tau_2^2 = 0.024$. When we did not perform any feature selection and only ran linear regression per each feature and iterated over all of them, on the same 5000 samples, we obtained $\tau_1^2 = 0.048$, and $\tau_2^2 = 0.030$.

3.5.3 Bootstrap and replication approach

We know that the value of τ_1^2 and τ_2^2 increases as the number of sample points increases due to reduction in variance of Gaussian posteriors. Based on this observation, we ran simulations where we replicated or bootstrapped the first 5000 samples in order to generate more data points. Tables 3.2, 3.3, 3.4 and 3.5 summarize the values of τ_1^2 and τ_2^2 in various scenarios where we replicated or bootstrapped 5000 samples in order to generate 20k, 40k, 60k and 80k data points. Each table contains empirical Bayes variances in 7 scenarios mentioned in details below.

1. "adlasso (all 1st) or (selected 1st)" refers to performing adaptive lasso and keeping all first order features or only the selected ones.

20k-replication	τ_1^2	τ_2^2	40k-replication	τ_1^2	τ_2^2
adlasso (all 1st)	0.553	-0.266	adlasso (all 1st)	0.843	0.230
adlasso (selected 1st)	0.323	-0.111	adlasso (selected 1st)	0.650	0.448
elastic net (all 1st)	-0.150	-0.588	elastic net (all 1st)	0.011	-0.257
elastic net (selected 1st)	-0.154	-0.586	elastic net (selected 1st)	0.009	-0.256
lasso (all 1st)	-0.121	-0.598	lasso (all 1st)	0.042	-0.276
lasso (selected 1st)	-0.154	-0.591	lasso (selected 1st)	0.009	-0.267
no feature selection	-0.173	-0.614	no feature selection	-0.012	-0.306

Table 3.2: 20k and 40k replication scenarios. Here "adlasso" refers to adaptive lasso.

2. "elastic net (all 1st) vs (selected 1st)" refers to performing elastic net and keeping all first order features or only the selected ones.

3. "lasso (all 1st) vs (selected 1st)" refers to performing lasso and keeping all first order features or only the selected ones.

4. "no feature selection": We replicate or bootstrap the data and run BLIP on the generated data set. We do not perform any feature selection.

In all of these cases, whenever we performed feature selection, we replicated/bootstrapped the data first and only used the selected second order features. We then ran BLIP on the generated data set. The empirical Bayes variances reported in these tables reveal that not having a convergent model can have a big effect on both τ_1^2 and τ_2^2 values. A model is called convergent if the Gaussian posterior distributions of the average weights have sufficiently small variances.

The objective for the next experiment was to investigate the effect of extra samples in computing empirical Bayes variances. A summary of our results are given in Table 3.6 and Table 3.7. We included the following in the experiment.

60k-replication	τ_1^2	τ_2^2	80k-replication	τ_1^2	τ_2^2
adlasso (all 1st)	1.038	0.694	adlasso (all 1st)	1.19	1.13
adlasso (selected 1st)	0.875	0.943	adlasso (selected 1st)	1.04	1.39
elastic net (all 1st)	0.114	0.050	elastic net (all 1st)	0.190	0.336
elastic net (selected 1st)	0.113	0.052	elastic net (selected 1st)	0.190	0.338
lasso (all 1st)	0.145	0.021	lasso (all 1st)	0.220	0.298
lasso (selected 1st)	0.113	0.031	lasso (selected 1st)	0.189	0.306
no feature selection	0.092	-0.021	no feature selection	0.169	0.245

Table 3.3: 60k, and 80k replication scenarios. Here "adlasso" refers to adapative lasso.

20k-bootstrapping	τ_1^2	τ_2^2	40k-bootstrapping	τ_1^2	τ_2^2
adlasso (all 1st)	0.551	-0.324	adlasso (all 1st)	0.852	0.241
adlasso (selected 1st)	0.214	-0.168	adlasso (selected 1st)	0.714	0.460
elastic net (all 1st)	-0.099	-0.587	elastic net (all 1st)	0.032	-0.250
elastic net (selected 1st)	-0.102	-0.586	elastic net (selected 1st)	0.027	-0.248
lasso (all 1st)	-0.073	-0.570	lasso (all 1st)	0.110	-0.241
lasso (selected 1st)	-0.093	-0.563	lasso (selected 1st)	0.075	-0.235
no feature selection	-0.159	-0.602	no feature selection	-0.012	-0.297

Table 3.4: 20k, and 40k bootstrapping scenarios. Here "adlasso" refers to adaptive lasso.

60k-bootstrapping	τ_1^2	τ_2^2
adlasso (all 1st)	1.04	0.759
adlasso (selected 1st)	0.875	1.01
elastic net (all 1st)	0.104	0.043
elastic net (selected 1st)	0.102	0.045
lasso (all 1st)	0.144	0.044
lasso (selected 1st)	0.141	0.046
no feature selection	0.092	-0.010

80k-bootstrapping	τ_1^2	τ_2^2
adlasso (all 1st)	1.16	1.17
adlasso (selected 1st)	1.04	1.42
elastic net (all 1st)	0.199	0.328
elastic net (selected 1st)	0.197	0.330
lasso (all 1st)	0.261	0.368
lasso (selected 1st)	0.257	0.370
no feature selection	0.179	0.263

Table 3.5: 60k, and 80k bootstrapping scenarios. Here "adlasso" refers to adaptive lasso.

1. We performed adaptive lasso with the full data set of 30,162 samples. We kept all the first order features and only the selected second order ones. We then ran BLIP on these features and calculated τ^2 values.

2. We replicated or bootstrapped the first 5000 samples to generate 30,162 data points. We then followed the process in number one to get τ^2 values.

Replication (30,162)	τ_1^2	τ_2^2
adaptive lasso (all 1st)	0.726	-0.005
adaptive lasso (selected 1st)	0.529	0.189
elastic net (all 1st)	-0.059	-0.427
elastic net (selected 1st)	-0.061	-0.425
lasso (all 1st)	-0.014	-0.429
lasso (selected 1st)	-0.043	-0.422

Bootstrap (30,162)	τ_1^2	τ_2^2
adaptive lasso (all 1st)	0.722	0.012
adaptive lasso (selected 1st)	0.537	0.202
elastic net (all 1st)	0.015	-0.419
elastic net (selected 1st)	0.258	-0.414
lasso (all 1st)	0.021	-0.412
lasso (selected 1st)	0.231	-0.403

Table 3.6: Empirical Bayes variances, τ_1^2 and τ_2^2, based on replication/bootstrapping of 5k samples to obtain 30,162 artificial data points (size of train data set). We then perform feature selection (adaptive lasso, elastic net and lasso) on the artificial data set, and fed it to BlipBayes to get τ^2 values.

Original data set (30,162)	τ_1^2	τ_2^2
adaptive lasso (all 1st)	0.636	-0.340
adaptive lasso (selected 1st)	1.21	-0.210
elastic net (all 1st)	-0.073	-0.712
elastic net (selected 1st)	-0.079	-0.709
lasso (all 1st)	0.030	-0.706
lasso (selected 1st)	-0.146	-0.689

Table 3.7: Empirical Bayes variances, τ_1^2 and τ_2^2, on all the original train data set of size 30,162 samples. We used the whole train data set and performed feature selection (adaptive lasso, elastic net, and lasso). We then fed the new data set to BlipBayes to obtain τ^2 values.

3.6 Prediction Accuracy

In this section, we report prediction accuracies for six experiments. We first describe the three scenarios compared in subsection 3.6.1. We then go over metrics used in reporting our prediction accuracies and report the results of our 6 experiments in subsection 3.6.2.

3.6.1 Scenarios Description

We are comparing three scenarios which are different only for the training on the first day as described below.

- Blip: We start with the traditional prior distribution as $N(0, 1)$ for all first and second order features and train the model on the first 5k samples.
- BlipBayes: In this scenario, we train the BLIP model on artificially built samples (based on bootstrapping of the first 5k samples) and compute τ_1^2 and τ_2^2. We then reset the prior for first-order features to $N(0, \tau_1^2)$ and for the second-order features to $N(0, \tau_2^2)$. We then run Blip on the first 5k samples using empirical Bayes priors.
- BlipTwice: Here, we run the BLIP model with a $N(0, 1)$ prior distribution on the same artificially generated samples as in the BlipBayes scenario. We then saved the model and trained it further on the first 5k samples.

We ran Blip, BlipBayes, and BlipTwice for 6 days and compared their prediction accuracies at the end of each day. We performed multiple experiments as reported in subsection 3.6.2 and section 3.7. The size of the train data on each day is 5027 samples. The reason we included BlipTwice was to make sure the comparison among Blip and BlipBayes was fair and consistent with the amount of data we fed into each model.

The metrics used to compare prediction accuracies for the aforementioned three scenarios are log loss and classification loss using a threshold value at 0.5. The log loss formula for

binary classification for N data points is given as

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \ln p_i + (1-y_i)\ln(1-p_i)].$$

3.6.2 Experiments: prediction accuracies

Experiments Description

First Experiment: In this experiment, we took 5k samples and bootstrapped the samples to obtain a 40k artificial data set and then performed regularized linear regression with adaptive lasso. We kept all first order features (13 of them) and only the selected second order features. The values of τ^2 as reported in Table 3.4 are $\tau_1^2 = 0.852$ and $\tau_2^2 = 0.241$. Figure 3.2 demonstrates the results of adaptive lasso feature selection. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.3 and reported in Table 3.8.

Second Experiment: In this experiment, we took 5k samples and bootstrapped the samples to obtain a 60k artificial data set and then performed regularized linear regression with adaptive lasso. We kept all first order features (13 of them) and only the selected second order features. The τ^2 values as reported in Table 3.5, are $\tau_1^2 = 1.04$ and $\tau_2^2 = 0.759$. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.4 and reported in Table 3.9.

Third Experiment: In this experiment, we took 5k samples and bootstrapped the samples to obtain an 80k artificial data set and then performed regularized linear regression with adaptive lasso. We kept all first order features (13 of them) and only the selected second order features (12 of them). The τ^2 values as reported in Table 3.5 are $\tau_1^2 = 1.16$ and $\tau_2^2 = 1.17$. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.5 and reported in Table 3.10.

Fourth Experiment: In this experiment, we took 5k samples and bootstrapped the samples to obtain a 40k artificial data set and then performed regularized linear regression with

adaptive lasso. We kept only selected features for both first and second order features. The number of selected first and second order features are 7 and 11 respectively. The τ^2 values as reported in Table 3.4 are $\tau_1^2 = 0.714$ and $\tau_2^2 = 0.460$. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.6 and reported in Table 3.11.

Fifth Experiment: In this experiment, we took 5k samples and bootstrapped the samples to obtain a 60k artificial data set and then performed regularized linear regression with adaptive lasso. We kept only the selected features for both first and second order features. The number of selected first order features is 8 and the number of selected second order features is 12. The τ^2 values as reported in Table 3.5 are $\tau_1^2 = 0.875$ and $\tau_2^2 = 1.01$. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.7 and reported in Table 3.12.

Sixth Experiment: In this experiment, we took 5k samples and bootstrapped them to obtain an 80k artificial data set and then performed regularized linear regression with adaptive lasso. We kept selected features for both first and second order features. The number of selected first order features is 8 and the number of selected second order features is 12. The τ^2 values as reported in Table 3.5 are $\tau_1^2 = 1.04$ and $\tau_2^2 = 1.42$. Log loss and classification loss (threshold at 0.5 level) values are plotted in Figure 3.8 and reported in Table 3.13.

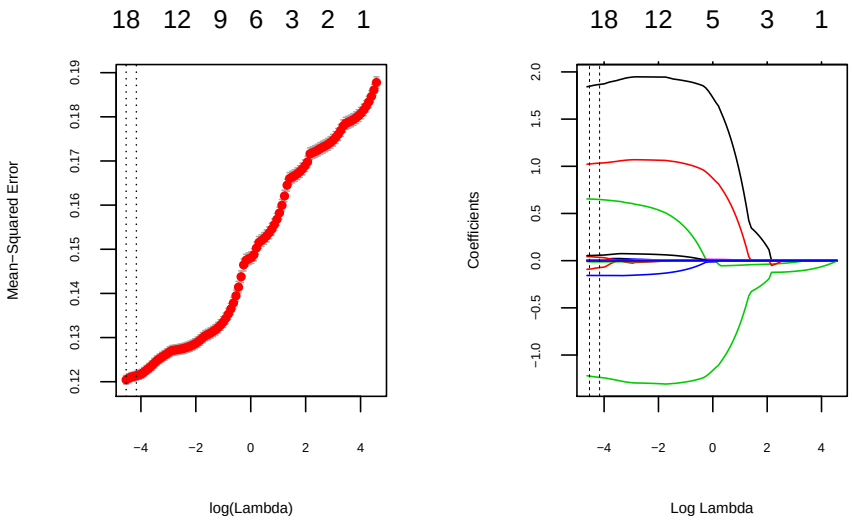

Figure 3.2: First Experiment: MSE and coefficients selected by adaptive lasso on a 40k built data set by bootstrapping of first 5k samples. We keep all 13 of the first order features, and the number of selected second order features by adaptive lasso is 11. The two dashed lines denote the log of $\lambda_{min} = -4.5378$ and $\lambda_{1se} = -4.1657$ obtained by 10-fold cross validation.

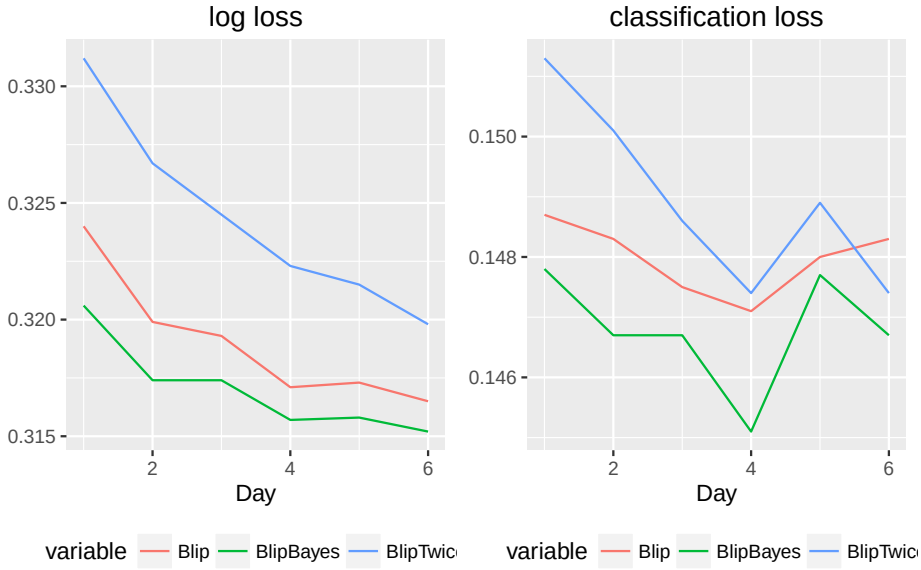

Figure 3.3: First Experiment: Log loss and 0/1 classification for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 0.852$ and $\tau_2^2 = 0.241$ respectively.

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3240	0.3206	0.3312	0.1487	0.1478	0.1513
Day 2	0.3199	0.3174	0.3267	0.1483	0.1467	0.1501
Day 3	0.3193	0.3174	0.3245	0.1475	0.1467	0.1486
Day 4	0.3171	0.3157	0.3223	0.1471	0.1451	0.1474
Day 5	0.3173	0.3158	0.3215	0.1480	0.1477	0.1489
Day 6	0.3165	0.3152	0.3198	0.1483	0.1467	0.1474

Table 3.8: First Experiment: Log loss and classification loss values for three scenarios; Blip, BlipBayes and BlipTwice. The features included in the data are based on adaptive lasso feature selection on a 40k artificially built data set (bootstrapped on 5k samples); keeping all 13 first-order features and only the selected second order features (11 of them). The empirical Bayes variances for first and second order features are $\tau_1^2 = 0.852$ and $\tau_2^2 = 0.241$ respectively. Size of the test set is 15060.

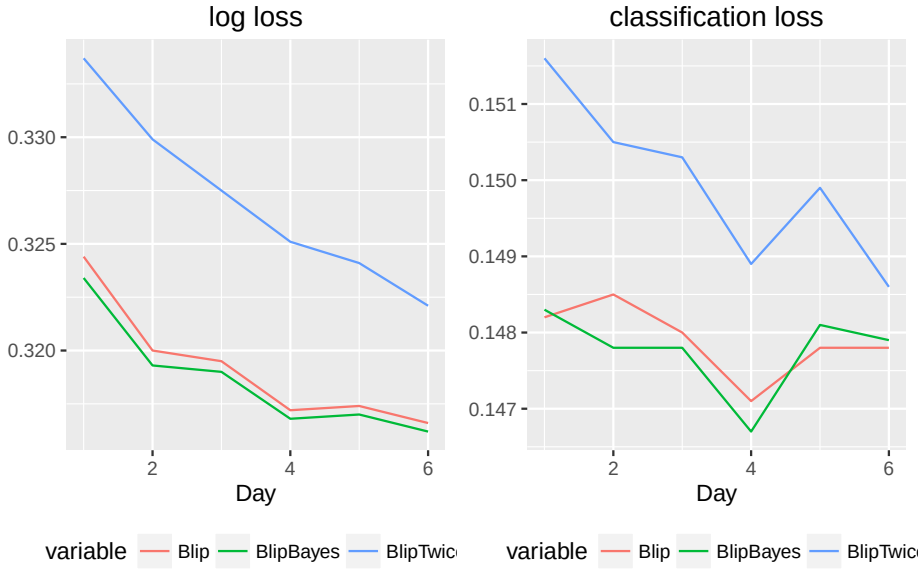

Figure 3.4: Second Experiment: Log loss and 0/1 classification loss (threshold at 0.5) for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 1.04$ and $\tau_2^2 = 0.759$ respectively.

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3244	0.3234	0.3337	0.1482	0.1483	0.1516
Day 2	0.3200	0.3193	0.3299	0.1485	0.1478	0.1505
Day 3	0.3195	0.3190	0.3275	0.1480	0.1478	0.1503
Day 4	0.3172	0.3168	0.3251	0.1471	0.1467	0.1489
Day 5	0.3174	0.3170	0.3241	0.1478	0.1481	0.1499
Day 6	0.3166	0.3162	0.3221	0.1478	0.1479	0.1486

Table 3.9: Second Experiment: Log loss and classification loss values for three scenarios; BLip, BlipBayes and BlipTwice. The features included in the data are based on adaptive lasso feature selection on a 60k artificially built data set (bootstrapped on 5k samples). We kept all 13 first order features and only the selected second order features (12 of them). The empirical Bayes variances for first and second order features are $\tau_1^2 = 1.04$ and $\tau_2^2 = 0.759$ respectively. Size of the test set is 15060.

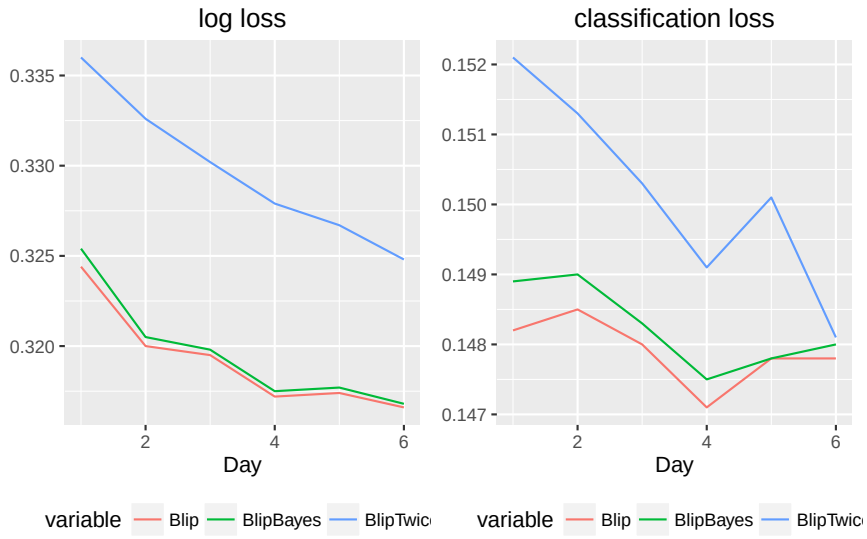

Figure 3.5: Third Experiment: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 1.16$ and $\tau_2^2 = 1.17$ respectively.

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3244	0.3254	0.3360	0.1482	0.1489	0.1521
Day 2	0.3200	0.3205	0.3326	0.1485	0.1490	0.1513
Day 3	0.3195	0.3198	0.3302	0.1480	0.1483	0.1503
Day 4	0.3172	0.3175	0.3279	0.1471	0.1475	0.1491
Day 5	0.3174	0.3177	0.3267	0.1478	0.1478	0.1501
Day 6	0.3166	0.3168	0.3248	0.1478	0.1480	0.1481

Table 3.10: Third Experiment: Log loss and classification loss values for three different scenarios; BLip, BlipBayes and BlipTwice. The features included in the data are based on adaptive lasso feature selection on an 80k artificially built data set (bootstrapped of 5k samples). We kept all 13 first order features and only the selected second order features (12 of them). The empirical Bayes variances for first and second order features respectively are $\tau_1^2 = 1.16$ and $\tau_2^2 = 1.17$. Size of the test set is 15060.

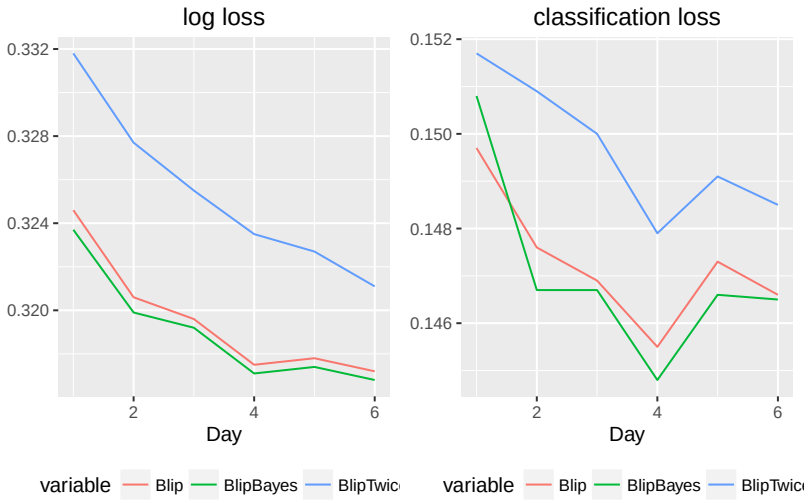

Figure 3.6: Fourth Experiment: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 0.714$ and $\tau_2^2 = 0.460$ respectively. We keep only the selected features for both first (7 of them) and second order ones (11 of them).

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3246	0.3237	0.3318	0.1497	0.1508	0.1517
Day 2	0.3206	0.3199	0.3277	0.1476	0.1467	0.1509
Day 3	0.3196	0.3192	0.3255	0.1469	0.1467	0.1500
Day 4	0.3175	0.3171	0.3235	0.1455	0.1448	0.1479
Day 5	0.3178	0.3174	0.3227	0.1473	0.1466	0.1491
Day 6	0.3172	0.3168	0.3211	0.1466	0.1465	0.1485

Table 3.11: Fourth Experiment: Log loss and classification loss values for three different scenarios BLip, BlipBayes, and BlipTwice. The features included in the data are based on adaptive lasso feature selection on a 40k artificially built data set (bootstrapped on 5k samples). We only keep the selected features for both first (7 features) and second order features (11 features). The size of the test set is 15,060. The empirical Bayes variances for both first and second order features are $\tau_1^2 = 0.714$ and $\tau_2^2 = 0.460$ respectively.

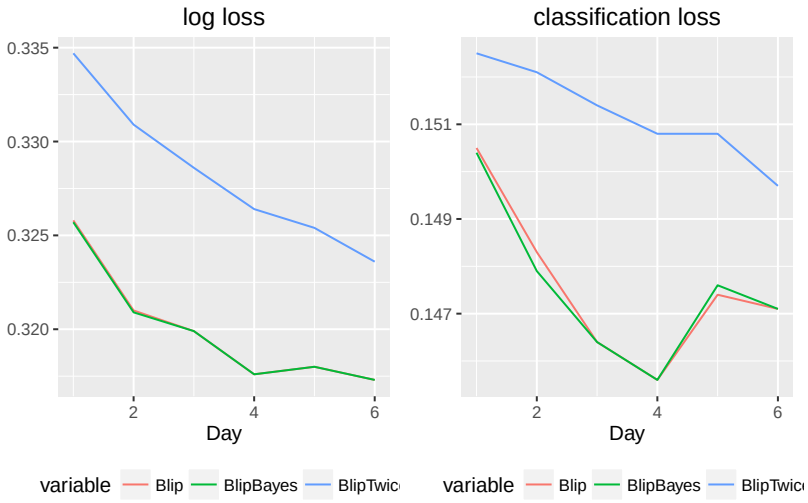

Figure 3.7: Fifth Experiment: Log loss and 0/1 classification loss (threshold at 0.5 level) for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 0.875$ and $\tau_2^2 = 1.01$ respectively. The number of selected first order features is 8 and the number of selected second order features is 12.

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3258	0.3257	0.3347	0.1505	0.1504	0.1525
Day 2	0.3210	0.3209	0.3309	0.1483	0.1479	0.1521
Day 3	0.3199	0.3199	0.3286	0.1464	0.1464	0.1514
Day 4	0.3176	0.3176	0.3264	0.1456	0.1456	0.1508
Day 5	0.3180	0.3180	0.3254	0.1474	0.1476	0.1508
Day 6	0.3173	0.3173	0.3236	0.1471	0.1471	0.1497

Table 3.12: Fifth Experiment: Log loss and classification loss (threshold at 0.5 level) values for three different scenarios BLip, BlipBayes and BlipTwice. The features included in the data are also based on adaptive lasso feature selection on a 60k artificially built data set (bootstrapped on 5k samples). We only keep the selected features for both first and second order features. The number of selected first order features is 8 and for second order features is 12. The empirical Bayes variances for first and second order features are $\tau_1^2 = 0.875$ and $\tau_2^2 = 1.01$ respectively. The size of the test set is 15,060.

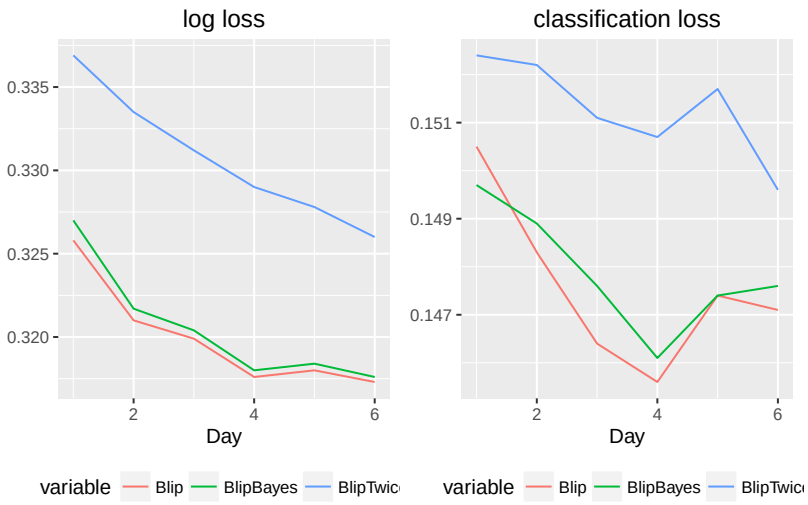

Figure 3.8: Sixth Experiment: Log loss and 0/1 classification loss (threshold at 0.5 level) for Blip, BlipBayes, and BlipTwice. The empirical Bayes variances for first and second order features are $\tau_1^2 = 1.04$ and $\tau_2^2 = 1.42$ respectively.

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3258	0.3270	0.3369	0.1505	0.1497	0.1524
Day 2	0.3210	0.3217	0.3335	0.1483	0.1489	0.1522
Day 3	0.3199	0.3204	0.3312	0.1464	0.1476	0.1511
Day 4	0.3176	0.3180	0.3290	0.1456	0.1461	0.1507
Day 5	0.3180	0.3184	0.3278	0.1474	0.1474	0.1517
Day 6	0.3173	0.3176	0.3260	0.1471	0.1476	0.1496

Table 3.13: Sixth Experiment: Log loss and classification loss (threshold at 0.5 level) values for three different scenarios BLip, BlipBayes and BlipTwice. The features included in the data are based on adaptive lasso feature selection on an 80k artificially built data set (bootstrapped on 5k samples). We keep the selected features for both first and second order features. The number of selected first order features is 8 and for second order features is 12. The empirical Bayes variances for first and second order features are $\tau_1^2 = 1.04$ and $\tau_2^2 = 1.42$ respectively. The size of the test set is 15060.

3.7 More Experiments

In this section, we compare prediction accuracies for Blip, BlipBayes, and BlipTwice in new scenarios. Specifically, in subsection 3.7.1 we reset the prior on the third day and in subsection 3.7.2 we perform experiments on 30 days and compute prediction accuracies.

3.7.1 Empirical Bayes Reset on 3rd Day

In this experiment, we divide the train data set into six chunks, each having 5,027 samples. We train the Blip model for three days (on about 15k samples) and compute empirical Bayes variances at the end of third day. We then reset the Gaussian weights to the empirical Bayes prior and re-train on all six chunks as in Blip. The prediction accuracies on days 3, 4, 5, and 6 for Blip and BlipBayes with prior reset after the first day and after the third day are plotted in Figure 3.9 and the values are reported in Table 3.14. Moreover, the empirical Bayes data set is the bootstrapped version of the first 15k samples after running adaptive lasso on it (keeping all first order features and only selected second order features). The values of the empirical Bayes variances are $\tau_1^2 = 0.862$ and $\tau_1^2 = 0.414$.

3.7.2 Empirical Bayes performance on 30 Days

In this subsection, we report the results of three experiments. In all, we generate data by bootstrapping the first 1000 samples, performing adaptive lasso on the generated data set, keeping all first order features and only the selected second order ones. We then run BLIP on the first day (1000 samples) and compute empirical Bayes variances and reset the prior for BlipBayes. In experiment 8, 9, and 10, we generated 40k, 15k, and 12k samples. Figures 3.10, 3.12, and 3.14 demonstrate the results of adaptive lasso feature selection. The log loss and classification loss values for experiments 8, 9, and 10 are presented in Figures 3.11, 3.13, and 3.15.

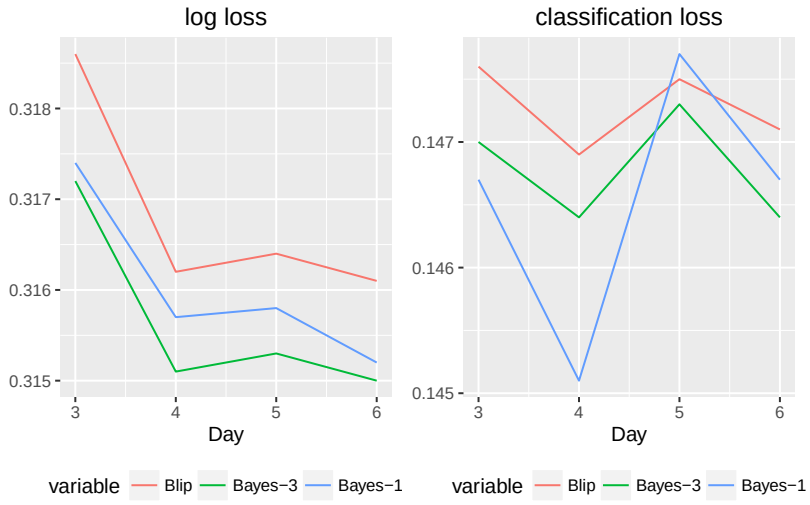

Figure 3.9: Prior Reset Experiment: Log loss and classification loss values after prior reset on day 1 (blue curve) and on day 3 (grren curve). The values of the empirical Bayes variances are $\tau_1^2 = 0.862$ and $\tau_2^2 = 0.414$ computed after the third day. The number of first order features included is 13 and the count of selected second order features is 11.

Loss/Day		Log loss		Classification loss
Model/Day	Blip	BlipBayes	Blip	BlipBayes
Day 3	0.3186	0.3172	0.1476	0.1470
Day 4	0.3162	0.3151	0.1469	0.1464
Day 5	0.3164	0.3153	0.1475	0.1473
Day 6	0.3161	0.3150	0.1471	0.1464

Table 3.14: Prior Reset Experiment: Log loss and classification loss (threshold at 0.5 level) values for BLip and BlipBayes. The empirical Bayes data set is the bootstrapped version of the first 15k samples after running adaptive lasso on it (keeping all first order features and only selected second order features). The values of the empirical Bayes variances are $\tau_1^2 = 0.862$ and $\tau_1^2 = 0.414$ computed on the third day of training. The size of the test set is 15,060.

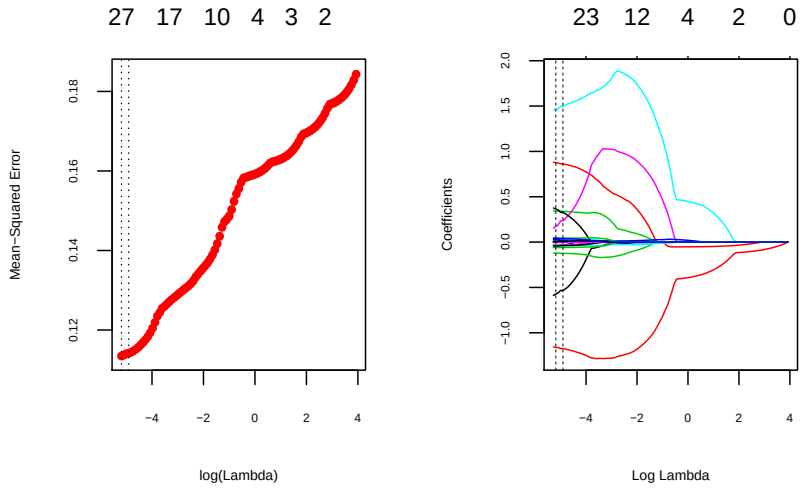

Figure 3.10: Experiment 8: MSE and coefficients selected by adaptive lasso on 40k built data set by bootstrapping of the first 1000 samples. Here we keep all 13 of the first order features and the number of selected second order features by adaptive lasso is 21. The two dashed lines denote the log of $\lambda_{min} = -5.1862$ and $\lambda_{1se} = -4.9071$ obtained by 10-fold cross validation.

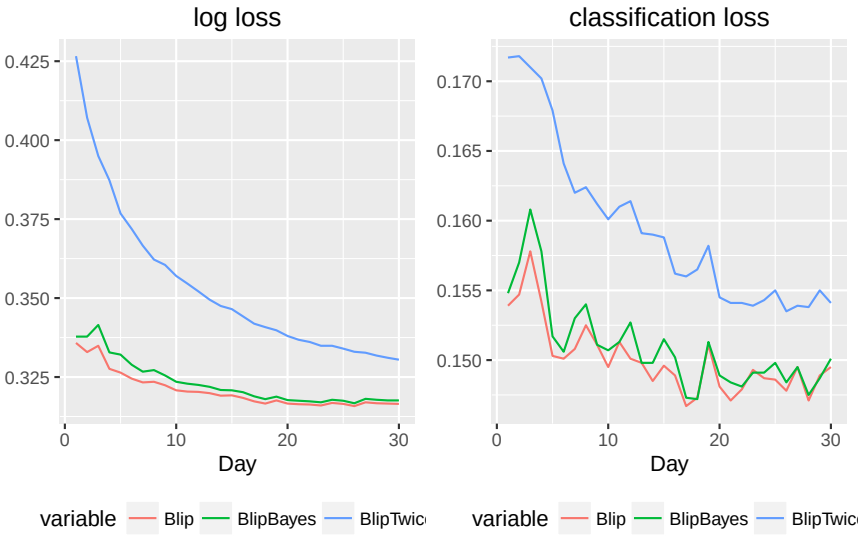

Figure 3.11: Experiment 8: Log loss and 0/1 classification for Blip, BlipBayes, and BlipTwice. The empirical Bayes variance for the first and second order features respectively are $\tau_1^2 = 1.84$ and $\tau_2^2 = 1.78$.

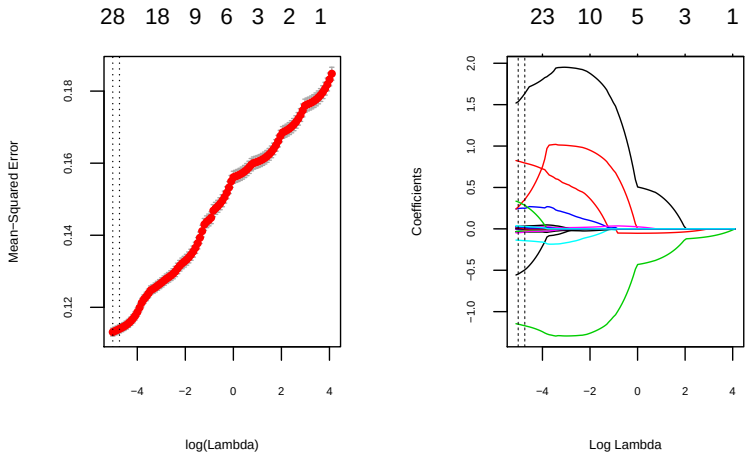

Figure 3.12: Experiment 9: MSE and coefficients selected by adaptive lasso on a 15k built data set by bootstrapping of the first 1000 samples. Here we keep all 13 of the first order features and the number of selected second order features by adaptive lasso is 20. The two dashed lines denote the log of $\lambda_{min} = -5.0204$ and $\lambda_{1se} = -4.7413$ obtained by 10-fold cross validation. The values of empirical Bayes variance are $\tau_1^2 = 0.939$ and $\tau_2^2 = 0.410$.

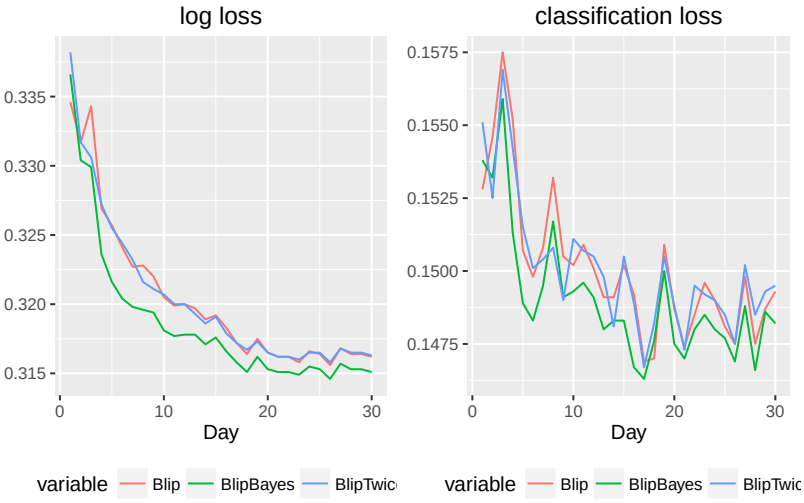

Figure 3.13: Experiment 9: Log loss and 0/1 classification for Blip, BlipBayes, and BlipTwice. The empirical Bayes variance for first and second order features respectively are $\tau_1^2 = 0.939$ and $\tau_2^2 = 0.402$.

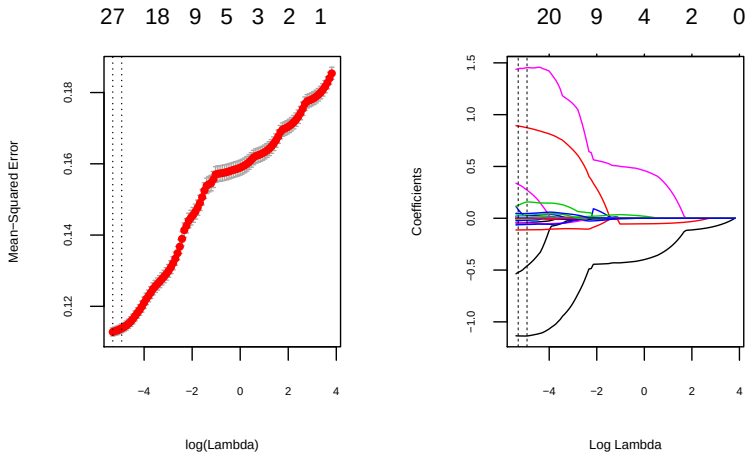

Figure 3.14: Experiment 10: MSE and coefficients selected by adaptive lasso on a 12k built data set by bootstrapping of the first 1000 samples. Here we keep all 13 of the first order features and the number of selected second order features by adaptive lasso is 21. The two dashed lines denote the log of $\lambda_{min} = -5.3047$ and $\lambda_{1se} = -4.9325$ obtained by 10-fold cross validation. The values of the empirical Bayes variances are $\tau_1^2 = 0.799$ and $\tau_2^2 = 0.132$.

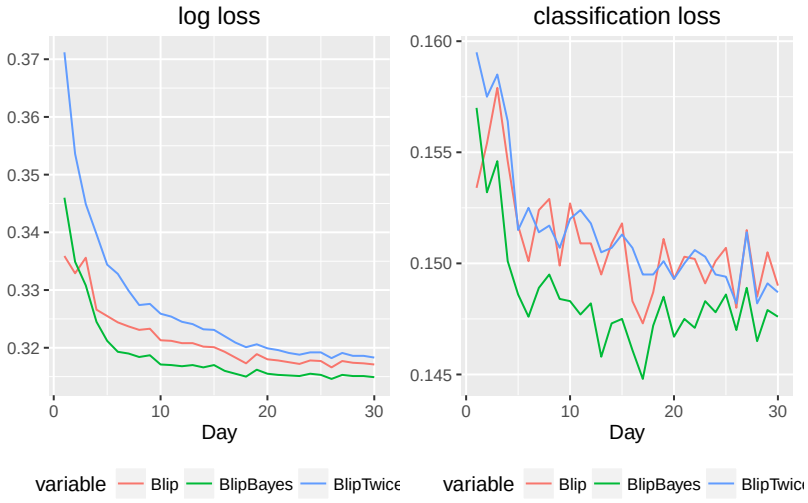

Figure 3.15: Experiment 10: Log loss and 0/1 classification for Blip, BlipBayes, and BlipTwice. The empirical Bayes variance for first and second order features respectively are $\tau_1^2 = 0.799$ and $\tau_2^2 = 0.132$.

3.7.3 3 Days experiment

In this subsection, we present experiment 11 which was performed over three days. We bootstrap 60k samples based on the first 10k data points and perform adaptive lasso; keeping all first order features and only the selected second order features. Empirical Bayes variances for first and second order features are $\tau_1^2 = 0.719$ and $\tau_2^2 = 0.275$ respectively. Figure 3.16 present log loss and classification loss and Table 3.15 report loss values.

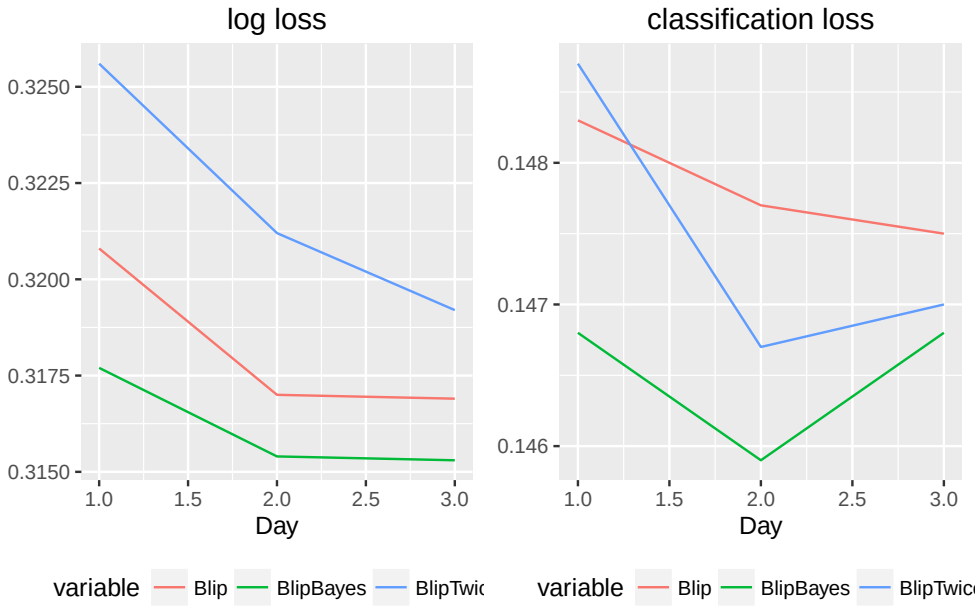

Figure 3.16: Experiment 11: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. Empirical Bayes variances for first and second order features are $\tau_1^2 = 0.719$ and $\tau_2^2 = 0.275$ respectively.

3.7.4 Extreme Runs

The objective of this subsection is to analyze the empirical Bayes performance in the case of non-optimal empirical Bayes prior variances. In the following experiments, we are taking 40k bootstrapped of the first 5k samples and then set the empirical Bayes prior to non-optimal

Loss/Day	Log loss			Classification loss		
Model/Day	Blip	BlipBayes	BlipTwice	Blip	BlipBayes	BlipTwice
Day 1	0.3208	0.3177	0.3256	0.1483	0.1468	0.1487
Day 2	0.3170	0.3154	0.3212	0.1477	0.1459	0.1467
Day 3	0.3169	0.3153	0.3192	0.1475	0.1468	0.1470

Table 3.15: Experiment 11: Log loss and classification loss for 3 days. Each day has about 10k samples and we bootstrap 60k samples based on the first 10k samples to compute τ^2 which are $\tau_1^2 = 0.719$ and $\tau_2^2 = 0.275$.

values. More specifically, we set the empirical Bayes variances to $\tau_1^2 = 5.0$ and $\tau_2^2 = 5.0$ in Figure 3.17, to $\tau_1^2 = 0.01$ and $\tau_2^2 = 0.01$ in Figure 3.18, and to $\tau_1^2 = 0.1$ and $\tau_2^2 = 0.1$ in Figure 3.19. The optimal empirical Bayes variances for first and second order features are $\tau_1^2 = 0.852$ and $\tau_2^2 = 0.241$ respectively. Figure 3.20 puts all the aforementioned figures in the same plot along with loss values for optimal empirical Bayes prior variances.

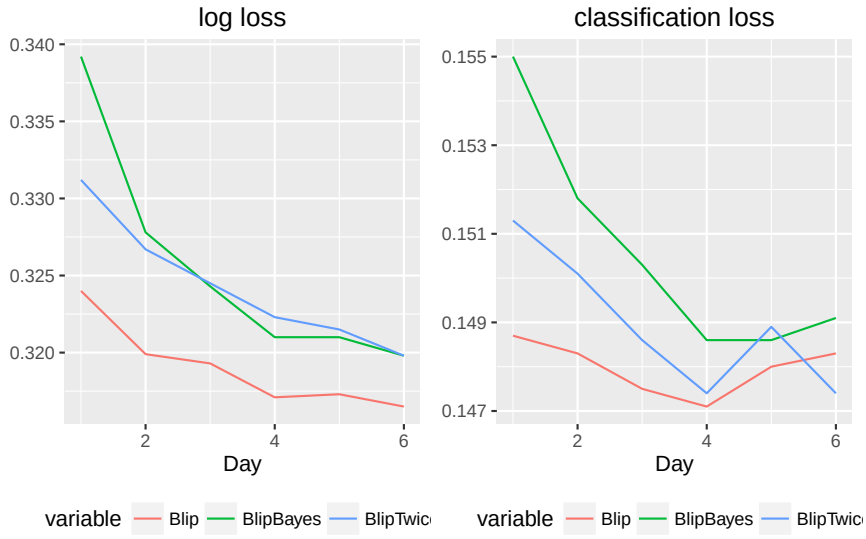

Figure 3.17: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. We set the empirical Bayes prior variances for first and second order features to their non-optimal values $\tau_1^2 = 5.0$ and $\tau_2^2 = 5.0$.

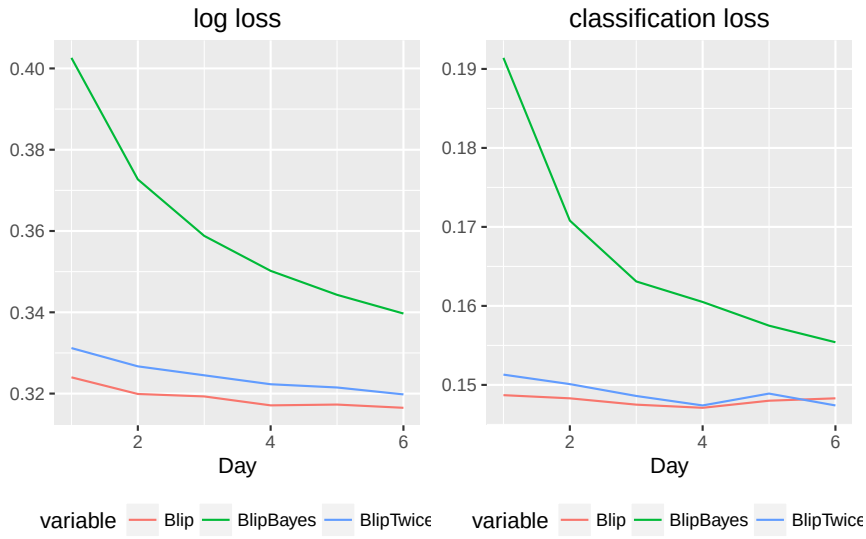

Figure 3.18: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. We set the empirical Bayes variances for first and second order features to their non-optimal values $\tau_1^2 = 0.01$ and $\tau_2^2 = 0.01$.

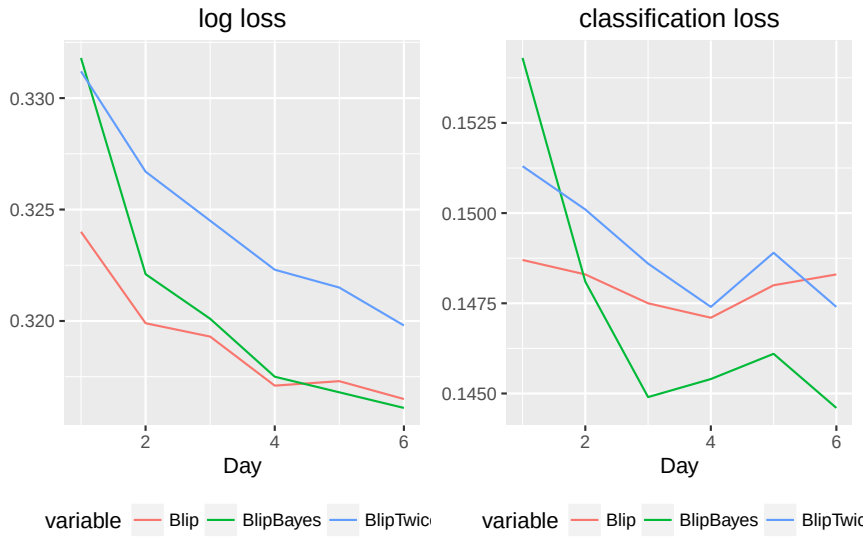

Figure 3.19: Log loss and 0/1 classification loss for Blip, BlipBayes, and BlipTwice. We set the empirical Bayes variances for first and second order features to their non-optimal values $\tau_1^2 = 0.1$ and $\tau_2^2 = 0.1$.

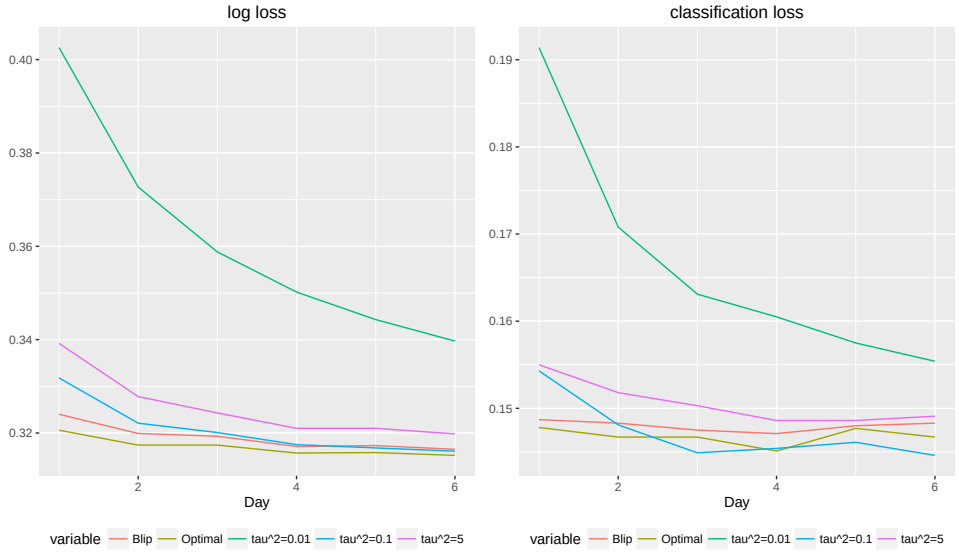

Figure 3.20: Log loss values for Blip and BlipBayes with different values of prior variances. The optimal empirical Bayes variances for first and second order features are $\tau_1^2 = 0.852$ and $\tau_2^2 = 0.241$ respectively. We also present log loss values for non-optimal empirical Bayes prior variances $\tau_1^2 = \tau_2^2 \in \{0.01, 0.1, 5\}$.

3.8 Future Directions

1. The BLIP model with the empirical Bayes prior is currently running in the production system of one of the largest retailers. The next step is to further understand/improve the BLIP Bayes model (BLIP with the empirical Bayes prior) on a much larger scale.
2. The idea of the empirical Bayes prior can be applied in transfer learning scenarios where a prior can be learned using a high volume of data, and be applied to scenarios where not enough data is available.
3. In the BLIP model and the parameter's update equations, independence between features is assumed (i.e. diagonal covariance matrix). As a future direction, we can investigate new parameter update equations using message passing where the covariance matrix is assumed to be non-diagonal.

BIBLIOGRAPHY

[1] Arian Aflaki, Pnina Feldman, and Robert Swinney. Choosing to be strategic: Implications of the endogenous adoption of forward-looking purchasing behavior on multiperiod pricing. 2016.

[2] Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, volume 23. JMLR Workshop and Conference Proceedings, 2012.

[3] Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In *AISTATS*, pages 99–107, 2013.

[4] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135, 2013.

[5] Victor F Araman and René Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations research*, 57(5):1169–1188, 2009.

[6] Victor F Araman and René Caldentey. Revenue management with incomplete demand information. *Wiley Encyclopedia of Operations Research and Management Science*, 2011.

[7] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[8] Yossi Aviv, Yuri Levin, and Mikhail Nediak. Counteracting strategic consumer behavior in dynamic pricing systems. In *Consumer-Driven Demand and Operations Management Models*, pages 323–352. Springer, 2009.

[9] Yossi Aviv and Amit Pazgal. A partially observed markov decision process for dynamic pricing. *Management Science*, 51(9):1400–1416, 2005.

[10] Yossi Aviv and Amit Pazgal. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, 10(3):339–359, 2008.

[11] Yossi Aviv and Gustavo Vulcano. Dynamic list pricing. In *The Oxford handbook of pricing management*. 2012.

[12] Yossi Aviv, Mingcheng Mike Wei, and Fuqiang Zhang. Responsive pricing of fashion

products: The effects of demand learning and strategic consumer behavior. Technical report, Working Paper, Washington University, 2015.

[13] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 207–216. IEEE, 2013.

[14] Donald A Berry and Bert Fristedt. *Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability)*, volume 12. Springer, 1985.

[15] David Besanko and Wayne L. Winston. Optimal price skimming by a monopolist facing rational concumers. *Management Science*, 36(5):555–567, 1990.

[16] Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

[17] Omar Besbes and Assaf Zeevi. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739, 2015.

[18] Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

[19] Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

[20] Jeremy I Bulow. Durable-goods monopolists. *Journal of political Economy*, 90(2):314–332, 1982.

[21] Gérard P Cachon and Robert Swinney. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science*, 55(3):497–511, 2009.

[22] Gérard P Cachon and Robert Swinney. The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science*, 57(4):778–795, 2011.

[23] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Advances in neural information processing systems*, pages 2249–2257, 2011.

[24] Minho Cho, Ming Fan, and Yong-Pin Zhou. Strategic consumer response to dynamic pricing of perishable products. In *Consumer-Driven Demand and Operations Management Models*, pages 435–458. Springer, 2009.

[25] Ronald H Coase. Durability and monopoly. *The Journal of Law and Economics*, 15(1):143–149, 1972.

[26] Eric Cope. Bayesian strategies for dynamic pricing in e-commerce. *Naval Research Logistics (NRL)*, 54(3):265–281, 2007.

[27] José Correa, Ricardo Montoya, and Charles Thraves. Contingent preannounced pricing policies with strategic consumers. *Operations Research*, 64(1):251–272, 2016.

[28] Arnoud V den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015.

[29] Arnoud V den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015.

[30] Arnoud V den Boer and Bert Zwart. Simultaneously learning and optimizing using controlled variance pricing. *Management science*, 60(3):770–783, 2013.

[31] Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5. Cambridge University Press, 2016.

[32] Wedad Elmaghraby, Altan Gülcü, and Pınar Keskinocak. Designing optimal preannounced markdowns in the presence of rational customers with multiunit demands. *Manufacturing & Service Operations Management*, 10(1):126–148, 2008.

[33] Serkan S Eren and Costis Maglaras. Monopoly pricing with limited demand information. *Journal of revenue and pricing management*, 9(1-2):23–48, 2010.

[34] Vivek F Farias and Benjamin Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

[35] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2015.

[36] Kris Johnson Ferreira, David Simchi-Levi, and He Wang. Online network revenue management using thompson sampling. 2016.

[37] Jérémie Gallien. Dynamic mechanism design for online commerce. *Operations Research*, 54(2):291–310, 2006.

[38] Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *COLT*, pages 359–376, 2011.

[39] Vishal Gaur and Marshall L Fisher. In-store experiments to determine the impact of price on sales. *Production and Operations Management*, 14(4):377–387, 2005.

[40] JC Gittins. Multi-armed bandit allocation indices. wiley-interscience series in systems and optimization. 1989.

[41] John Gittins, Kevin Glazebrook, and Richard Weber. *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

[42] Thore Graepel, Joaquin Q Candela, Thomas Borchert, and Ralf Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 13–20, 2010.

[43] Ole-Christoffer Granmo. Solving two-armed bernoulli bandit problems using a bayesian learning automaton. *International Journal of Intelligent Computing and Cybernetics*, 3(2):207–234, 2010.

[44] J Michael Harrison, N Bora Keskin, and Assaf Zeevi. Dynamic pricing with an unknown linear demand model: asymptotically optimal semi-myopic policies. *Access: http: faculty-gsb. atanford. edu/harrison/hkz-2. pdf*, 2011.

[45] J Michael Harrison, N Bora Keskin, and Assaf Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586, 2012.

[46] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On bayesian upper confidence bounds for bandit problems. In *AISTATS*, pages 592–600, 2012.

[47] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*, pages 199–213. Springer, 2012.

[48] N Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

[49] Jawad Khan. What is dynamic pricing & how does it affect ecommerce?, 2017.

[50] Nathaniel Korda, Emilie Kaufmann, and Rémi Munos. Thompson sampling for 1-dimensional exponential family bandits. In *Advances in Neural Information Processing Systems*, pages 1448–1456, 2013.

[51] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[52] Yuri Levin, Jeff McGill, and Mikhail Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management science*, 55(1):32–46, 2009.

[53] Yuri Levin, Jeff McGill, and Mikhail Nediak. Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. *Production and Operations Management*, 19(1):40–60, 2010.

[54] Tatsiana Levina, Yuri Levin, Jeff McGill, and Mikhail Nediak. Dynamic pricing with online learning and strategic consumers: An application of the aggregating algorithm. *Operations Research*, 57(2):327–341, 2009.

[55] Jun Li, Nelson Granados, and Serguei Netessine. Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114–2137, 2014.

[56] Qian Liu and Garrett J Van Ryzin. Strategic capacity rationing to induce early purchases. *Management Science*, 54(6):1115–1131, 2008.

[57] Choncé Maddox. The importance of pricing for profit, 2017.

[58] Odalric-Ambrym Maillard, Rémi Munos, Gilles Stoltz, et al. A finite-time analysis of multi-armed bandits problems with kullback-leibler divergences. In *COLT*, pages 497–514, 2011.

[59] Vincent Mak, Amnon Rapoport, Eyran J Gisches, and Jiaojie Han. Purchasing scarce products under dynamic pricing: An experimental investigation. *Manufacturing & Service Operations Management*, 16(3):425–438, 2014.

[60] Benedict C May, Nathan Korda, Anthony Lee, and David S Leslie. Optimistic bayesian sampling in contextual-bandit problems. *Journal of Machine Learning Research*, 13(Jun):2069–2106, 2012.

[61] Benedict C May and David S Leslie. Simulation studies in optimistic bayesian sampling in contextual-bandit problems. In *Technical Report 11: 02. Statistics Group, Department of Mathematics*. University of Bristol, 2011.

[62] Adam J Mersereau and Dan Zhang. Markdown pricing with unknown fraction of strategic customers. *Manufacturing & Service Operations Management*, 14(3):355–370, 2012.

[63] Thomas P Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[64] Rafi Mohammed. Use pricing strategy to boost sales, 2012.

[65] Harikesh Nair. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics*, 5(3):239–292, 2007.

[66] Serguei Netessine and Christopher S Tang. *Consumer-driven demand and operations management models: a systematic study of information-technology-enabled sales mechanisms*, volume 131. Springer Science & Business Media, 2009.

[67] Jack Nicas. Now prices can change from minute to minute, 2015.

[68] U.S. Department of Commerce. U.s. census bureau news, 2018.

[69] Nikolay Osadchiy and Elliot Bendoly. Are consumers really strategic? implications from an experimental study. *Implications from an Experimental Study (October 2015)*, 2015.

[70] Özalp Özer and Robert Phillips. *The Oxford handbook of pricing management*. Oxford University Press, 2012.

[71] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[72] Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

[73] Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[74] Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of thompson sampling. *The Journal of Machine Learning Research*, 17(1):2442–2471, 2016.

[75] Schumpeter. Flexible figures: A growing number of companies are using "dynamic" pricing, 2016.

[76] Steven L Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658, 2010.

[77] Zuo-Jun Max Shen and Xuanming Su. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and operations management*, 16(6):713–728, 2007.

[78] Gonca P Soysal and Lakshman Krishnamurthi. Demand dynamics in the seasonal goods industry: An empirical analysis. *Marketing Science*, 31(2):293–316, 2012.

[79] Stax. Customer decision making criteria and the importance ofprice, October 2016. [Online; posted 24-October-2016].

[80] Nancy L Stokey. Rational expectations and durable goods pricing. *The Bell Journal of Economics*, pages 112–128, 1981.

[81] Xuanming Su. Intertemporal pricing with strategic customer behavior. *Management Science*, 53(5):726–741, 2007.

[82] Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[83] Liang Tang, Romer Rosales, Ajit Singh, and Deepak Agarwal. Automatic ad format selection via contextual bandits. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1587–1594. ACM, 2013.

[84] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[85] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[86] Volodimir G Vovk. Aggregating strategies. In *Proc. Third Workshop on Computational Learning Theory*, pages 371–383. Morgan Kaufmann, 1990.

Appendix A

KNOWN FORMULAS APPLIED

Fact 1. Chernoff-Hoeffding bounds

Let $X_1, ..., X_n$ be random variables with common range [0,1] with the property that $E[X_t|X_1, ..., X_{t-1}] = \mu$. Let $S_n = X_1 + ... + X_n$. Then, for all $a \geq 0$,

$$P_r(S_n \geq n\mu + a) \leq e^{\frac{-2a^2}{n}},$$
$$P_r(S_n \leq n\mu - a) \leq e^{\frac{-2a^2}{n}}.$$

Fact 2. Pinsker's Inequality

$$KL(p, q) \geq 2(p - q)^2.$$

where $KL(p, q)$, denotes the KL-divergence between two distributions p, and q.

For two Bernoulli distributions, $p \sim Bernoulli(\mu_1)$, and $q \sim Bernoulli(\mu_2)$,

$$KL(p, q) = \mu_1 \ln \frac{\mu_1}{\mu_2} + (1 - \mu_1) \ln \frac{1 - \mu_1}{1 - \mu_2}$$

and for two Gaussian distributions $p \sim (\mu_1, \sigma_1^2)$, and $q \sim (\mu_2, \sigma_2^2)$,

$$KL(p, q) = \ln \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

In the case of two Gaussian distributions with equal variances,

$$KL(p, q) = \frac{(\mu_1 - \mu_2)^2}{2}.$$

Appendix B

POSTERIOR DISTRIBUTION

B.1 Beta Posterior

In this appendix, we give more details of the Beta distribution, and the Bayesian framework for prior and posterior distributions. In the Bayesian framework, the problem is to infer a distribution for a parameter θ given observed data x. In Bayesian statistics, they first assume a prior distribution $p(\theta)$ on parameter θ. Then they calculate the likelihood function $p(x|\theta)$ using observed data x. Finally, the posterior distribution of parameter θ is computed as:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$$

If the posterior distribution $(p(\theta|x))$ is in the same family as the prior distribution $(p(\theta))$, then the prior and posterior distributions are called conjugate distributions. The prior is called conjugate prior for the likelihood function. All distributions from the exponential family have conjugate priors.

In our setting, the retailer assumes a prior distribution on $d_k \sim Beta(\alpha, \beta)$ where d_k is the true customer purchase probability under price p_k which is unknown to the retailer. Also, suppose after $t-1$ periods, price p_k is offered $N_{k,(t-1)}$ times. Out of all these, the retailer observes $R_{k,(t-1)}$ customer purchases under price p_k. Therefore, the likelihood of this observation is given as

$$p(R_{k,(t-1)}, N_{k,(t-1)}|d_k = x) = C_{R_{k,(t-1)},N_{k,(t-1)}}\ x^{R_{k,(t-1)}}\ (1-x)^{N_{k,(t-1)}-R_{k,(t-1)}} \tag{B.1}$$

We also have the Beta prior on d_k, therefore,

$$p(d_k = x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{B.2}$$

where

$$\begin{aligned} B(\alpha,\beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \end{aligned}$$

$B(\alpha,\beta)$ is called the Beta function. By applying the Bayesian framework, we derive the posterior distribution of d_k as

$$\begin{aligned} p(d_k = x|R_k, N_k) &= \frac{p(R_k, N_k|d_k = x)p(d_k = x)}{\int_0^1 p(R_k, N_k|d_k = x)p(d_k = x)\, dx} \\ &= \frac{x^{R_k+\alpha-1}(1-x)^{N_k-R_k+\beta-1}}{B(\alpha+R_k, \beta+N_k-R_k)} \\ &\sim Beta(\alpha+R_k, \beta+N_k-R_k) \end{aligned}$$

B.2 Gaussian Posterior

Here, we prove that given prior $\tilde{\mu} \sim N(\hat{\mu}(t), v^2 B_i^{-1}(t))$, the posterior distribution follows $\tilde{\mu} \sim N(\hat{\mu}(t+1), v^2 B_i^{-1}(t+1))$. Here are the underline assumptions:

- $r_i(t)|p_i, \mu \sim N(p_i\mu_i, v^2)$
- demand under price p_i has $N(\mu_i, (\frac{v}{p_i})^2)$
- $r_i(t)$ is Gaussians.

Assume at time t, arm i is pulled and revenue $r_i(t)$ is observed. The update steps of our proposed algorithm give

$$B_i(t+1) = B_i(t) + p_i^2,\ f_i = f_i + r_i(t),\ \hat{\mu}(t) = B_i^{-1}(t) f_i$$

We know that

$$p_r(\tilde{\mu}|r_i(t)) \propto p_r(r_i(t)|\tilde{\mu}) p_r(\tilde{\mu})$$

Since $r_i(t)|\tilde{\mu} \sim N(p_i\tilde{\mu}, v^2)$, and $\tilde{\mu} \sim N(\hat{\mu}(t), v^2 B_i^{-1}(t))$, we get

$$\begin{aligned} p_r(\tilde{\mu}|r_i(t)) &\propto exp\Big\{\frac{-1}{2v^2}(r_i(t) - p_i\tilde{\mu})^2 + \frac{-1}{2v^2 B_i^{-1}(t)}(\tilde{\mu} - \hat{\mu}_i(t))^2\Big\} \\ &\propto exp\Big\{\frac{-1}{2v^2}\Big[r_i(t)^2 - 2r_i(t)p_i\tilde{\mu} + p_i^2\tilde{\mu}^2 + B_i(t)\tilde{\mu}^2 - 2\tilde{\mu}\hat{\mu}_i(t)B_i(t) + B_i(t)\hat{\mu}_i(t)^2\Big]\Big\} \end{aligned}$$