Title
Topics in Modeling Uncertainty with Learning

Author
Jain, Ankit

Publication Date
2010

Topics in Modeling Uncertainty with Learning

By

Ankit Jain

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Industrial Engineering and Operations Research

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Associate Professor Andrew E. B. Lim, Co-Chair
Emeritus Professor J. George Shanthikumar, Co-Chair
Associate Professor Martin Wainwright

Spring 2010

Topics in Modeling Uncertainty with Learning

Abstract

Topics in Modeling Uncertainty with Learning

by

Ankit Jain

Doctor of Philosophy in Industrial Engineering and Operations Research

University of California, Berkeley

Associate Professor Andrew E. B. Lim, Co-Chair
Emeritus Professor J. George Shanthikumar, Co-Chair

It is fair to say that in many real world decision problems the underlying models cannot be accurately represented. Uncertainty in a model may arise due to lack of sufficient data to calibrate the model, non-stationarity, or due to wrong subjective assumptions. Hence optimization in presence of model uncertainty is a very important issue. In the last few decades, there has been a lot of work on finding robust solutions to model uncertainty in operations research. With advances in the field of convex optimization, many robust optimization problems are efficiently solvable. Still there are many challenges and open questions related to model uncertainty, specially when learning is also involved. In this thesis, we study the following challenges and problems related to model uncertainty with learning:

First, defining and computing a robust solution to problems with model uncertainty is a challenging task, specially in dynamic optimization problems. Many dynamic problems are tractable only because the structure of the solutions can be analyzed and therefore the dimension of a solution space can be reduced. It is therefore important to design and study robust equivalents of dynamic optimization problems and analyze the structure of robust solutions.

Second, in many situations the need for robust solution arises because of lack of sufficient data to calibrate model parameters. It is important to study the properties of traditional robust solutions. Are these robust solutions good for decision making in long run as compared to non-robust solutions? If not, is it possible to design alternative approaches?

Third, it is not possible have robustness to model uncertainty if the attention is restricted to a wrong model class. One may use a relatively model free learning and decision making methodology such as reinforcement learning, which can achieve an optimal solution asymptotically. However, these approaches may take a long time to learn and therefore it is important to look at non-parametric methodologies which can achieve good small sample performance.

The main contribution of this thesis can be summarized as follows:

1. We study an infinite-time queuing control system, where both arrival and departure rates can be controlled. We consider the case when arrival and departure processes can not be accurately modeled. We prove that a threshold policy is optimal under a max-min robust optimization objective when the uncertainty in the processes is characterized by a novel notion of relative entropy. The new notion of relative entropy accounts for different levels of modeling errors in arrival and departure processes.

2. We perform numerical tests to study the performance of traditional robust optimization solutions with learning using past few data points. It is shown that the performance of a robust solution may even be worse than a classical point estimate based non-robust solution. We introduce the notion of generalized operational statistics that guarantees a better solution than a classical solution over a set of uncertain parameters, while incorporating subjective prior information. We apply operational statistics approach to mean-variance portfolio optimization problem with uncertain mean returns. We show that the operational statistics portfolio problem can be efficiently solvable by reformulating it as a semi-definite program. Various extensions are discussed and numerical experiments are done to show the efficacy of the solution.

3. We introduce objective operational learning, a new non-parametric approach that incorporates structural information to improve small sample performance. We show how structural and objective information can be incorporated in the objective operational learning algorithm. We apply the algorithm to an inventory control problem with demand dependent on inventory level and prove convergence.

To my family and friends

Contents

List of Figures

List of Tables

Acknowledgments

My sincere and heartfelt gratitude to all those who have helped and inspired me during my doctoral studies.

First and foremost, I like to thank my PhD adviser and dissertation committee chair Prof. Andrew Lim for all the support he has given to me over the years. Due to his friendly and caring attitude it was great fun working with him. His dedication and enthusiasm have always inspired me to work harder.

I am very grateful to my co-adviser and dissertation committee member Prof. George Shanthikumar for his valuable ideas and insights. I thank him for patiently listening to me during our meetings and presentations. Without his vision and continual encouragement, the thesis work would not have been possible.

I am grateful to my dissertation committee member Prof. Martin Wainwright for his valuable comments during my dissertation workshop and also for his critical comments that improved the written presentation of my thesis.

I thank all the past and present members of my research group, Thaisiri Watewai, Peng Li, Lian Yu, Gah-Yi Vahn, Sea Chen, Poomyos Fook, Vivek Ramamurthy and Huanning Cai, for their helpful remarks during various group presentations.

I also thank all the instructors and GSIs of the courses I have taken. I like to specially thank Prof. David Aldous, Prof. Jim Pitman, Dr. Shankar Bhamidi, Prof. Michael Jordan, Prof. Peter Bartlett, Prof. Alper Atamtürk and Prof. Xin Guo for enriching the technical skills required for the dissertation work.

I express thanks to my friends and colleagues, Dr. Vishnu Narayanan, Dr. Rahul Tandra, Dr. Ambuj Tewari, Dr. Ismail Ceylan, Dr. Ismail Onur Filiz, Engin Alper, Erick Moreno-Centeno and Sehzad Wadalawala, for many useful discussions and remarks. I also like to thank the IEOR social club, specially Kory Hedman and Stella So, for organizing various social activities and making my stay fun.

I am grateful to all the members of the IEOR administrative staff, specially Mike Campbell, Anayancy Paz and Jay Sparks, for their administrative support.

I thank my mentors during my internships at IBM T. J. Watson research, Dr. Cathy Xia and Dr. Zhen Liu, for their valuable comments on my work, and for providing me opportunities to apply my research in broader areas.

I express my gratitude to all the attendees of my presentations at IEOR student seminars, INFORMS annual meetings and IBM research center for attentively listening to me and providing valuable feedback.

I thank all my friends and roommates at Berkeley, specially Puneet Bhargava, Gautam Gupta, Pankaj Kalra, Anurag Gupta and Shariq Rizvi, for making my stay at Berkeley a pleasant experience.

I express my sincere gratitude to all the researchers on whose work I am able to build upon and write this dissertation.

I thank my parents and all other family members for their unconditional love, support and encouragement. I am very grateful to my brother Ankur Jain for always encouraging and motivating me whenever I was feeling low.

Last but not the least, a very special thanks to Shrutivandana Sharma, without her support it would not have been possible to complete the dissertation.

Chapter 1

Introduction

In 1961, Daniel Ellsberg described in his work [Ell61], what is commonly known as Ellsberg Paradox. An example of the Ellsberg paradox is as follows: there are two urns containing red and black balls, from one of which a ball will be drawn at random. Let R_2 denotes the choice of betting on red ball from urn 2. Given the choice one will receive \$1, if the ball drawn is red and \$0, if the ball drawn is black. We define the choices R_1, B_2, B_1 and the associated rewards in a similar way. Now suppose we have the following information: urn 2 contains 100 red and black balls but in a ratio that is unknown and urn 1 contains 50 red and 50 black balls. Given this information people are asked to draw their preferences.

It has been observed that most of the people have the following preference:

$$R_1 \backsimeq B_1 \succ B_2 \backsimeq R_2, \tag{1.1}$$

i.e., people are indifferent between R_1 and B_1 but prefer B_1 or R_1 over B_2 or R_2. An observer applying the basic rules of probability and utility theory would infer tentatively that a subject regards the event 'black from urn 1' as more likely than 'black from urn 2'. She would also infer that 'red from urn 1' is preferred over 'red from urn 2'. Since she cannot

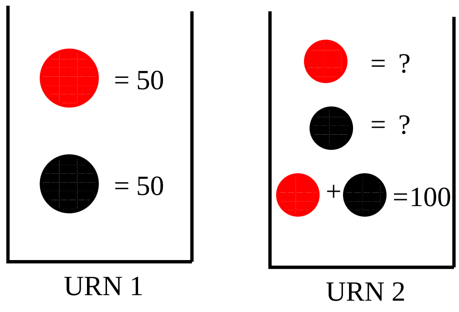

Figure 1.1: Ellsberg Paradox

conclude that 'red from urn 1' is more likely than 'red from urn 2' and at the same time 'not-red from urn 1' is more likely than 'not-red from urn 2', this behavior is inconsistent with the essential properties of probability relationships. There is no probability measure on the balls in the urn which supports the preferences described in (1.1).

A probable explanation of people's behavior in the above example is as follows: in case of urn 2, the subject has too little information to form a probability distribution (prior) on the number of red balls and black balls. Hence she considers a set of possible priors, and being uncertainty/ambiguity averse she calculates the minimal expected utility over all priors in her subjective uncertainty set while evaluating a bet. To explain the Ellsberg paradox using this logic, one may consider the extreme case in which the decision maker takes into account all possible priors over red and black balls in urn 2. In that case the minimal expected utility of the choice R_2 or B_2 is 0 as there is a prior which corresponds to all red balls or all black balls. On the other hand the minimal expected utility of the choice R_1 or B_1 is 50 as the exact prior is known. This explains the behavior observed in (1.1). This type of observed behavior in people's preferences is also called *ambiguity aversion*.

The notion of ambiguity or model uncertainty as separate from risk (that can be characterized by a probability distribution) was defined as early as 1921 by Frank Knight [Kni21]. Over the years ambiguity averse modeling has been studied in many optimization settings by researchers in the field of operations research (OR). One of the earliest papers in inventory control that considered ambiguity averse modeling was by Scarf [H58] in 1958. He used a min-max objective to find the optimal inventory control policy with unknown demand distribution, assuming only the precise knowledge of first two moments of the distribution. Gallego and Moon [GM94] extended the model of Scarf to several other cases such as inventory control with fixed ordering cost. Advances in the field of convex optimization [Wri97] in last three decades made it possible to solve ambiguity averse or robust problems, specially in the field of deterministic optimization, where the parameters (such as product ordering cost in an inventory control) of a decision problem are assumed to be uncertain (as opposed to being random). Typically these uncertain parameters are assumed to lie in a closed convex region like an ellipsoid or intervals. If an optimization problem with known parameters is given by:

$$\begin{aligned} &\max_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{(}a)) \\ &\text{subject to} \\ &\qquad g(\boldsymbol{x}, \boldsymbol{b}) \geq 0, \end{aligned} \tag{1.2}$$

then a robust version of the problem is:

$$\begin{aligned} &\max_{\boldsymbol{x}} \min_{\boldsymbol{a}\in\mathcal{A}, \boldsymbol{b}\in\mathcal{B}} f(\boldsymbol{x}, \boldsymbol{a}) \\ &\text{subject to} \\ &\qquad g(\boldsymbol{x}, \boldsymbol{b}) \geq 0, \qquad \forall b \in \mathcal{B}, \end{aligned} \tag{1.3}$$

where $\boldsymbol{x}$ is the decision variable and $\boldsymbol{a}, \boldsymbol{b}$ are parameters of the system. $\mathcal{A}$ and $\mathcal{B}$ are parameter uncertainty sets in the robust problem for $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively.

For many classes of functions f and g and uncertainty sets, including linear functions and ellipsoidal uncertainty sets, the robust problem is a "nice" convex program that can be efficiently solved. Some of the works in operations research along this line can be found in Ben-Tal and Nemirovski [BTN98, BTN99, BTN00], El Ghaoui and Lebret [EL97], Bertsimas and Sim [BS04], Bertsimas and Theile [BTar], El Ghaoui, Oks and Oustry [EOO03], Goldfarb and Iyengar [GI03]. In place of 'max-min', alternative objectives such as 'min-max regret' or 'competitive ratio' are used too. We study different objectives and their effect on optimization problems in Chapter 3.

Despite the advances, there has been comparatively little research in OR on robust or ambiguity averse dynamic optimization problems, where state of the world changes with time, mainly because of hardness of resulting dynamic programs. Nilim and El Ghaoui [Nl05] explored robust stochastic dynamic programs where state of the world evolves according to an uncertain probability transition matrix. Their work is closely related to earlier work in economics by Epstein and Wang [EW94] and Chen and Epstein [CE02]. Lim and Shanthikumar [LS07] considered a dynamic pricing problem where uncertainty in probability distribution of future states is characterized by a set of probability distributions which are at a certain "distance" to a nominal probability measure characterized by relative entropy.

In Chapter 2, we consider a dynamic optimization problem of controlling a single stage queuing system where arrivals and departures are modeled by point processes with stochastic intensities. An arrival incurs a cost while a departure earns a revenue. The objective is to maximize the profit by controlling the intensities subject to capacity limits and holding costs. When the stochastic model for arrival and departure processes are completely known, then a threshold policy is known to be optimal. We prove that a threshold policy is optimal under a max-min robust model, when the uncertainty in the processes is characterized by relative entropy. Our model generalizes the standard notion of relative entropy to account for different levels of model uncertainty in arrival and departure processes. Despite the criticism of max-min model (see Chapter 3) for being too conservative or being too sensitive with respect to the size of uncertainty set, max-min optimization remains an important tool in robustness and sensitivity analysis. First, it provides a class of policies parametrized by the size of uncertainty set, which a decision maker can choose from. Second, it is an important tool to analyze the effect of uncertainty on the decision. For example, in the queuing problem in Chapter 2, there are problem instances for which the controls are not affected by uncertainty in arrivals or departures.

Coming back to the discussion on the Ellsberg paradox and ambiguity modeling, one should be careful in applying the same concepts when there is repeated decision making and learning is involved. For example, if one is asked to play the same game as described by the Ellsberg paradox repeatedly, one can choose a red ball and a black ball alternatively from urn 2 and get the same utility as someone choosing a black or red ball from urn 1. Typically in many operations research problems, the uncertainty is due to lack of sufficient data to learn the parameters of the underlying distribution, non-stationarity in the stochastic process or wrong model assumptions. In classical modeling one tends to disregard the uncertainty in estimates of parameters or estimated distribution. It is also common to use the uncertainty sets estimated from the data, such as confidence intervals or ellipsoids, in robust optimization problems. it is not clear that if the decisions are made repeatedly, then the long run performance of robust models would be better than the classical ones. It is also important to know how much sensitive the robust models are to the size of uncertainty sets. These issues are explored in detail in Chapter 3.

Even when one is reasonably sure that the true parameter lies in an uncertainty set, it can not be guaranteed that the robust policy would outperform the classical policy. Ideally, instead of doing optimization and estimation separately, we want a mapping of past data to a policy which is optimal in some way, or is at least better than the classical policy in expected sense. This in theory may be achieved by defining the negative of objective function of the problem as a risk function and looking for an estimate which is uniformly better than the classical estimate. Unfortunately for many objective functions and distributions of underlying stochastic process a uniformly better estimate is impossible or hard to find.

In Chapter 4, we use a novel approach called *Operational Statistics* which aims to improve on a classical policy (in long run or expected sense) over a set of parameters. This is achieved by explicitly constraining the policy to be better than the classical policy over the set. The operational statistics formulation also incorporates subjective information (which may not necessarily be derived from data) about the underlying parameter. Thus, an operational statistics approach would strive to improve on the classical estimation based policy over the uncertainty set while also incorporating subjective belief about the underlying parameter of the stochastic process.

In Chapter 5, we apply the operational statistics approach to a mean-variance portfolio optimization problem with uncertain mean returns of stocks. Given mean return of stocks and covariance matrix, the objective of a mean-variance portfolio optimization problem introduced by Markowitz [Mar52] is to find the proportion of different stocks in the portfolio by maximizing a quadratic utility, which is equal to the average return of portfolio - variance of the portfolio multiplied by a risk aversion constant. Estimate of means from limited number of samples is known to be bad and the policy which replaces actual mean vector with sample mean vector is nown to perform bad out of sample. We show that our operational statistics mean variance portfolio optimization problem can be reformulated as a semi-definite program and thus can be solved reasonably efficiently. We show the connection of our approach with norm constrained portfolio optimization approach and discuss various extensions and numerical experiments.

Another important issue in repeated decision making, which involves repeated estimation and optimization steps, is the effect of optimization step on estimation, particularly if the model assumptions are wrong. It has been observed in revenue management problems [CdMK06] that using inaccurate model assumptions may lead to progressively worse estimates and in turn worse per step revenue. One simple example of such a phenomenon is the case of censored demand. Suppose someone is observing demand for a particular product by observing how many units of that particular product are sold on a retail store shelf. That person does not realize that the demand she observes in a particular period is the minimum of the actual demand and number of units on the shelf. In a stochastic demand setting the optimal number of units to order or place on shelf (see Chapter 3) is a particular quantile of the demand distribution. So if use quantiles of empirical distribution constructed from sales data, it may so happen that we may progressively have worse quantiles, and thus in turn have progressively worse “optimal” order quantity. This results in a per step revenue function that spiral down to zero. In general such a situation may arise when demand is dependent on existing inventory or in many pricing problems.

In presence of inaccurate model assumptions, one may be tempted to use a relatively model free approach like reinforcement learning [SB98] or multi-arm bandits [ACBF02]. However, there is more structural information available than what is typically used in a multi-arm bandit algorithm. For example, given the demand in a particular period and order quantity we know the exact functional form of profit function. Ignoring these structural information may lead to poor small sample performance. In Chapter 6 we introduce a new approach called *Objective Operational Learning* which utilizes this information efficiently. We apply the approach to an inventory control problem with demand dependent on inventory level. We show the comparison of objective operational learning approach to non-parametric regression and prove asymptotic convergence.

Chapter 2

Application of Dynamic Robust Optimization in Queueing Control

We consider a general single-stage queuing system, in which the input (arrival) and output (service completion) processes are modeled by point processes with dynamically controlled stochastic intensities. An entering job incurs a cost, $\tilde{c}$, and a job completion produces revenue, $\tilde{p}$. In addition there is a holding cost which is linearly proportional to the number of jobs in the system at a given time. The problem is to dynamically control both the input and output intensities so as to maximize discounted profit.

Problems of this type have been studied for example by Chen and Yao [CY90], where it is shown that a threshold policy for both the input and output processes is optimal under the assumption that the stochastic model for arrival and departure processes is accurate and known. (See the papers [Li88], [Sti85] and [Ser81] for similar results). In many applications, however, arrival and departure intensities can not be accurately modeled due to complexities of the real-world system or lack of sufficient calibration data. This raises natural questions including (i) what is the impact of model uncertainty on the "optimal" operating policies for the system, and (ii) are threshold policies still "optimal"? We account for model errors by formulating a max-min robust control version of this problem in which model uncertainty is incorporated using the notion of relative entropy. Within this framework, we show that threshold policy is optimal for the robust control problem, and study the impact of the level of model uncertainty on the optimal threshold level.

While the use of relative entropy to account for model uncertainty in stochastic optimization problems has a relatively long history ([PJD00], [LS07], [HSTW06] and [PMR96]), one feature of our work which departs from the standard approach is that we generalize the standard notion of relative entropy in order to allow for different levels of model uncertainty for the arrival as well as the departure processes (see also Lim, Shanthikumar and Watewai [LSW09] for similar ideas in the context of dynamic pricing). Aside from being realistic–for example, it is likely to be the case that the system operator is substantially more knowledgeable about the service system he/she in controlling (since it is internal) than the customer arrival process, which is typically much more complicated and subject to many external factors–this also allows us to study (say) the impact of the level of model uncertainty in the arrival process on the service control policy.

The outline of this chapter is as follows. In Section 2.1 we recall the model from Chen and Yao [CY90] and formulate the robust version of this problem. The robust version involves an extension of the notion of discounted relative entropy from Hansen, Sargent, Turmuhambetova and Williams [HSTW06] in order to handle different levels of model uncertainty for the arrival and departure processes. Dynamic programming equations for the robust control problem are derived in Section 2.2, and the impact of the level of model uncertainty on the threshold control levels is studied in Section 2.3.

2.1 Model Formulation

In this section we first introduce the standard model which is similar to [CY90] before formulating the robust model in section 2.1.2. The robust model extends the notion of discounted relative entropy from [HSTW06] in order to handle different level of uncertainties in arrival and departure rates.

2.1.1 Nominal Model

Consider a single-stage queuing system as shown in Fig. 2.1. Let X_t be the state of the system that denotes the number of jobs in process at time t. X_t takes values on nonnegative integers and is of the form

$$X_t = x_0 + A_t - D_t, \tag{2.1}$$

where $x_0 \geq 0$ is the state at time $t = 0$ of the system and A_t and D_t are the arrival and departure processes respectively. A_t and D_t denote the cumulative number of arrivals and departures until time t.

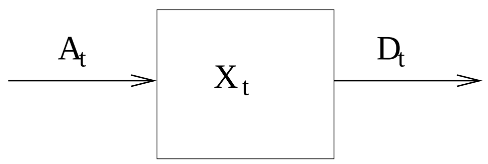

Figure 2.1: Queuing System

We assume that A_t and D_t are simple point processes. Let $\mathcal{F}_t$ be the sigma field generated by X_t, i.e., $\mathcal{F}_t = \sigma(X_s, s \leq t)$. Also let A_t and D_t admit $\mathcal{F}_t$ predictable intensities β_t and α_t.

The rates α_t and β_t are subjected to the following capacity constraints

$$\begin{aligned} &0 \le \beta_t \le y, \qquad \forall t \ge 0, \qquad \text{and} \\ &0 \le \alpha_t \le z, \qquad \forall t \ge 0. \end{aligned} \tag{2.2}$$

If there is no ambiguity in the arrival or departure process, i.e., if we can exactly control the arrival and departure intensities, then our objective is to find a control $u = \{\beta_t, \alpha_t, t \ge 0\}$ to maximize the following discounted value function:

$$V(x_0, u) = \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t}(\tilde{p}dD_t - \tilde{c}dA_t - hX_t dt). \tag{2.3}$$

Here $\mathbb{E}_{x_0}$ denotes the conditional expectation given $X_0 = x_0$, δ is the discount factor, $\tilde{p}$ is the revenue obtained by selling one unit of output, $\tilde{c}$ is the cost of acquiring one unit of input and h is the unit holding cost for work-in-process inventory. Substituting X_t from (2.1) in (2.3) we obtain:

$$V(x_0, u) = \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t} \left(\left(\tilde{p} + \frac{h}{\delta} \right) dD_t - \left(\tilde{c} + \frac{h}{\delta} \right) dA_t \right) - \frac{hx_0}{\delta}. \tag{2.4}$$

Defining $p = \tilde{p} + \frac{h}{\delta}$ and $c = \tilde{c} + \frac{h}{\delta}$ we have

$$V(x_0, u) = \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t} (pdD_t - cdA_t) - \frac{hx_0}{\delta}. \tag{2.5}$$

We can drop the last term in (2.5) for the purpose of finding optimal control as it is a constant term. From the definition of stochastic intensity [Bre81]

$$\begin{aligned} \mathbb{E}_{x_0} \int_0^\infty cdA_t &= \mathbb{E}_{x_0} \int_0^\infty c\beta_t dt, \\ \mathbb{E}_{x_0} \int_0^\infty pdD_t &= \mathbb{E}_{x_0} \int_0^\infty p\alpha_t dt. \end{aligned} \tag{2.6}$$

Rewriting the value function in (2.5) using (2.6) and dropping the constant term we have

$$V(x_0, u) = \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t} (pdD_t - cdA_t) = \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t} (p\alpha_t - c\beta_t)\, dt. \tag{2.7}$$

The problem formulation with unambiguous arrival rate is:

$$\max_u V(x_0, u) = \max_u \mathbb{E}_{x_0} \int_0^\infty e^{-\delta t} (p\alpha_t - c\beta_t)\, dt. \tag{2.8}$$

2.1.2 Robust Model

Let $(\Omega, \mathcal{F}_t, \mathcal{F})$ be the underlying measurable space for arrival and departure processes, A_t and D_t respectively. A_t and D_t are counting processes and admit intensities. A complete specification of intensity λ_t of the process A_t and of intensity μ_t of the process D_t induces a measure P over $\mathcal{F}$. The nominal model is based on the assumption that the decision maker is able to set arrival and departure intensities precisely subject to capacity constraints. The objective then is to find (λ_t, μ_t) which are optimal.

In reality the real-world intensity processes are unlikely to be (λ_t, μ_t). For example, the arrival rate, λ_t, might be a function of the price an arriving customer pays for the service being offered while μ_t could depend on the number of workers assigned to the customer in service, and the assumption in the nominal model is the decision maker knows the exact relationship between pricing decisions and the arrival rate λ_t, as well as the number of workers assigned and the departure rate μ_t, so that the arrival and departure rates can be set to the precise values that the decision maker desires. In practice, the relationship between the pricing decision and λ_t and also the number of assigned workers and the service rate μ_t may be difficult to characterize. The arrival intensity might be a complicated non-stationary function of the price and also of other factors such as amount of advertising. This makes it impossible to precisely calibrate intensities.

More generally we have a situation where the decision maker on the basis of her model thinks she is setting the arrival and departure rates at levels (λ_t, μ_t) but in reality the rates might be something different (say (β_t, α_t)). Our objective in this section is to incorporate the possibility of such model uncertainty into the formulation of the problem.

Suppose the real-world $\mathcal{F}_t$-predictable intensity processes β_t and α_t induces a measure Q over $\mathcal{F}$. We assume that the real-world intensity processes, while not known accurately, satisfy certain minimal conditions with respect to the intensity processes λ_t and μ_t, which are precisely known to the decision maker. Let P_t and Q_t be restrictions of P and Q respectively to $\mathcal{F}_t$. In particular we assume that for all t, Q_t is absolutely continuous with respect to P_t, i.e.,

$$P_t(A) = 0 \Rightarrow Q_t(A) = 0 \qquad \forall A \in \mathcal{F}_t.$$

The distribution Q is said to be absolutely continuous over finite intervals with respect to P if Q_t is absolutely continuous with respect to P_t for all t. This definition of absolute continuity captures the idea that two models are impossible to distinguish with certainty over a finite interval([HSTW06]).

Let $\{\gamma_t, t \geq 0\}$ be a stochastic process such that for every t, γ_t is Radon-Nikodym derivative [Dur03] of Q_t with respect to P_t. γ_t is a positive martingale and is adapted to filtration $\mathcal{F}_t$. It follows from [Jac79] that there are F_t-predictable processes κ_t and η_t such that:

$$\gamma_t = \exp\left(\int_0^t (\ln(\kappa_s)dA_s + \ln(\eta_s)dD_s) + \int_0^t ((1-\kappa_s)\lambda_s + (1-\eta_s)\mu_s)\, ds\right) \tag{2.9}$$

The following result is a version of the Girsanov Theorem for point processes as stated in Bremaud [Bre81].

Theorem 1 (Girsanov Theorem). *Let A_t and D_t be $\mathcal{F}_t$-adapted point processes with F_t-predictable intensities λ_t and μ_t respectively under the probability measure P. Suppose that γ_t is a positive F_t-martingale under P and that the Radon-Nikodym density of Q_t with respect to P_t is given by*

$$\frac{dQ_t}{dP_t} = \gamma_t = \exp\left(\int_0^t (\ln(\kappa_s)dA_s + \ln(\eta_s)dD_s) + \int_0^t ((1-\kappa_s)\lambda_s + (1-\eta_s)\mu_s)\, ds\right), \quad (2.10)$$

then A_t and D_t are $\mathcal{F}_t$-adapted point processes with intensities $\beta_t = \kappa_t\lambda_t$ and $\alpha_t = \eta_t\mu_t$ respectively under Q.

Theorem 1 allows us to parameterize the real-world model $Q = (\beta_t, \alpha_t, t \geq 0)$ through the processes κ_t and η_t.

2.1.3 Relative Entropy

Relative entropy or KL divergence is a measure of difference between two probability measures. Here we use a weaker notion, called *Discounted Relative Entropy* [HSTW06] to measure the discrepancy between two measures over an infinite horizon.

The weaker notion requires that the two measure being compared put positive probability on all of the same events, except tail events. The discounted relative entropy is defined as:

$$\tilde{\mathbb{R}}(Q|P) = \delta \int_0^\infty exp(-\delta t) \left(\int \ln\left(\frac{dQ_t}{dP_t}\right) dQ_t\right) dt, \quad (2.11)$$

where $\frac{dQ_t}{dP_t}$ is the Radon-Nikodym derivative of Q_t with respect to P_t.

This measure of relative entropy is convex in Q as shown in [HSTW06]. It should be noted here even if the discounted measure of entropy is finite the standard relative entropy measure of distance between P and Q can be infinite, i.e., it allows:

$$\int \log\left(\frac{dQ}{dP}\right) dQ = +\infty \quad (2.12)$$

If (2.12) holds but discounted relative entropy (2.11) is finite, then it means that a statistician would be able to distinguish between the probability measures P and Q with a continuous record of data on an infinite interval while it is impossible to do so by recording a finite length time interval data. As an example if under P the arrival rate is constant λ and under Q the arrival rate is constant β, $\beta \neq \lambda$, then relative entropy of P and Q is infinite but the discounted relative entropy between Q and P is finite.

Returning to our discussion on point processes, it follows from Theorem 1 that our measure of discounted relative entropy (2.11) transforms into:

$$\begin{aligned}\tilde{\mathbb{R}}(Q|P) &= \delta \int_0^\infty e^{-\delta t} \left(\int \ln \frac{dQ_t}{dP_t} dQ_t \right) dt \\ &= \delta \int_0^\infty e^{-\delta t} \left(\int_0^t (\lambda_s(\kappa_s \ln \kappa_s + 1 - \kappa_s) + \mu_s(\eta_s \ln \eta_s + 1 - \eta_s))\, ds \right) dt \\ &= \delta \int_0^\infty (\lambda_s(\kappa_s \ln \kappa_s + 1 - \kappa_s) + \mu_s(\eta_s \ln \eta_s + 1 - \eta_s))\, ds \left(\int_s^\infty e^{-\delta t} dt \right) \\ &= \int_0^\infty e^{-\delta s} \lambda_s(\kappa_s \ln \kappa_s + 1 - \kappa_s) ds + \int_0^\infty e^{-\delta s} \mu_s(\eta_s \ln \eta_s + 1 - \eta_s) ds.\end{aligned} \tag{2.13}$$

where the third equality is justified by Fubini's theorem [Dur03] as the integrand is positive. The first term $\tilde{\mathbb{R}}_1(Q|P) = \int_0^\infty e^{-\delta s} \lambda_s(\kappa_s \ln \kappa_s + 1 - \kappa_s)ds$ can be interpreted as measure of ambiguity in the arrival process. Similarly the second term $\tilde{\mathbb{R}}_2(Q|P) = \int_0^\infty e^{-\delta s} \mu_s(\eta_s \ln \eta_s + 1 - \eta_s)ds$ measures the ambiguity in the departure process.

Our robust control problem corresponding to (2.8) is as follows:

$$\begin{aligned}&\max_{u \in U} \min_Q \mathbb{E}_Q \left[\int_0^\infty e^{-\delta t} (p\alpha_t - c\beta_t dt) \right] \\ &\text{subject to: } \tilde{\mathbb{R}}(Q|P) \le \eta.\end{aligned} \tag{2.14}$$

Here the control is $u = \{\lambda_t, \mu_t, \lambda_t \le y, \mu_t \le z, t \ge 0\}$.

The robust control problem is a two-player game between 'nature' and decision maker. Given the control u, nature chooses a "worst-case" measure Q from the class of measures defined by the convex discounted relative entropy constraint. The constant $\eta \ge 0$ is a measure of our confidence in the nominal measure P and restricts the amount that Q (or the real-world intensity processes β_t and α_t) can deviate from P (resp. λ_t and μ_t). A large value of η allows Q to deviate further from our nominal probability measure P while a small value of η is chosen when we have a high degree of confidence in our nominal model. Putting $\eta = 0$ reduces the robust control problem to a standard one.

Alternatively, we may consider the following problem:

$$\max_{u \in U} \min_Q \left(\mathbb{E}_Q \left[\int_0^\infty e^{-\delta t} (p\mu_t - c\beta_t dt) \right] + \theta \tilde{\mathbb{R}}(Q|P) \right). \tag{2.15}$$

The constant $\theta > 0$ may be seen as the Lagrange multiplier for the relative entropy constraint in (2.14) and solving (2.14) is equivalent to solving (2.15) for an appropriate choice of θ. Alternatively, the parameter θ can represents our confidence in the nominal model. A large value of θ denotes high confidence in the model as the penalty of deviation from the model is large.

Note that the discounted relative entropy in (2.13) is the sum of two terms. The terms individually can be interpreted as measure of uncertainties in arrival and departure processes respectively. In formulation (2.15) as both the terms in discounted relative entropy expansion are multiplied by the same constant θ, the confidence levels in arrival and departure processes are assumed to be the same. If we have reason to believe in varying levels of confidence in arrival and departure processes the formulation (2.15) can be modified as:

$$\max_{u\in U}\min_{Q}\left(\mathbb{E}_Q\left[\int_0^\infty e^{-\delta t}\left(p\alpha_t - c\beta_t dt\right)\right] + \theta_A\tilde{\mathbb{R}}_1(Q|P) + \theta_D\tilde{\mathbb{R}}_2(Q|P)\right). \tag{2.16}$$

where θ_A and θ_D denotes the confidence in arrival and departure processes respectively . Hence Model (2.16) differs from standard robust model (2.15) that assumes the same level of uncertainty for all parts of the model.

Substituting the value of discounted relative entropy for point processes from (2.13) to (2.15) our robust formulation is:

$$\begin{aligned}\max_{u\in U}\min_{\kappa,\eta}\mathbb{E}_{x_0}\Big[\int_0^\infty e^{-\delta t}\big(&p\eta_s\mu_s - c\kappa_s\lambda_s + \theta_A\lambda_s(1-\kappa_s+\kappa_s\ln\kappa_s)\\ &+\theta_D\mu_s(1-\eta_s+\eta_s\ln\eta_s)\big)ds\Big].\end{aligned} \tag{2.17}$$

2.2 Characterization of Optimal Policy

Suppose we first restrict ourselves to the policies which are Markov in the state (the number of items that are currently in service). In other words, we can replace λ_t and μ_t by $\lambda(X_t)$ and $\mu(X_t)$ respectively. Further assume that nature is restricted to choose among a set of Markovian policy only, i.e., κ and η are only functions of X. In this case the formulation (2.17) reduces to:

$$\begin{aligned}\max_{\lambda,\mu}\min_{\kappa}\mathbb{E}_{x_0}\Big[\int_0^\infty e^{-\delta t}\big(&p\eta(X_s)\mu(X_s) - c\kappa(X_s)\lambda(X_s) + \theta_A\lambda(X_s)(1-\kappa(X_s)\\ &+\kappa(X_s)\ln\kappa(X_s)) + \theta_D\mu(X_s)(1-\eta(X_s)+\eta(X_s)\ln\eta(X_s))\big)ds\Big].\end{aligned} \tag{2.18}$$

The Hamiltonian-Jacobi-Bellman (HJB) equation corresponding to the above formulation is:

$$\begin{aligned}\delta V(x) = \max_{\lambda(x),\mu(x)}\min_{\kappa(x),\eta(x)}\Big[&\lambda(x)\kappa(x)(-c+\theta_A(-1+\ln\kappa(x))+\Delta V(x)) + \lambda(x)\theta_A\\ &+\mu(x)\eta(x)(p+\theta_D(-1+\ln\eta(x)) - \Delta V(x-1)) + \mu(x)\theta_D\Big],\end{aligned} \tag{2.19}$$

where

$$\Delta V(x) = V(x+1) - V(x), \tag{2.20}$$

and

$$V(x) = \max_{\lambda,\mu} \min_{\kappa,\eta} \mathbb{E}_x \Big[\int_0^\infty e^{-\delta t} \big(p\mu(X_s) - c\kappa(X_s)\lambda(X_s) + \theta_A \lambda(X_s)(1-\kappa(X_s) + \kappa(X_s)\ln\kappa(X_s)) + \theta_D \mu(X_s)(1-\eta(X_s)+\eta(X_s)\ln\eta(X_s)) \big) ds \Big]. \tag{2.21}$$

The solution of the (unconstrained convex) inner minimization (with respect to κ and η) problem in (2.19) is characterized by the first order conditions and yields the following

$$\kappa^*(x) = \exp\left(-\frac{1}{\theta_A}(\Delta V(x) - c) \right), \qquad \text{and}$$
$$\eta^*(x) = \exp\left(-\frac{1}{\theta_D}(p - \Delta V(x-1)) \right). \tag{2.22}$$

Substituting back the value of κ^* and $\eta^*(x)$ from (2.22) to (2.19), we obtain the following after some manipulation:

$$\delta V(x) = \max_{\lambda(x),\mu(x)} \Big[\theta_A \lambda(x) \Big(1 - \exp\Big(-\frac{1}{\theta_A}(\Delta V(x) - c) \Big) \Big) + \theta_D \mu(x) \left(1 - \exp(-\frac{1}{\theta_D}(p - \Delta V(x-1)) \right) \Big]. \tag{2.23}$$

As the above equation is linear in $\lambda(x)$ and $\mu(x)$, we obtain the following characterization of the optimal policy

$$\lambda^*(x) = \begin{cases} y & \text{if } \Delta V(x) \geq c \\ 0 & \text{otherwise.} \end{cases} \tag{2.24}$$

$$\mu^*(x) = \begin{cases} z & \text{if } \Delta V(x-1) \leq p, \qquad x \geq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.25}$$

This proves that optimal policy would either allow arrivals at full force or not to allow arrivals at all. The same structure holds for production. We either produce at full force or do not produce at all. In order to guarantee that the optimal policy is threshold we need to prove the existence of a number b such that:

$$\begin{cases} c \leq \Delta V(x) \leq p & \text{for } x \leq b \\ \Delta V(x) < c & \text{for } x > b. \end{cases} \tag{2.26}$$

As it does not make sense to stop the production if there is a positive inventory due to discounting and the holding cost, it is obvious that the optimal output policy should be of the following form:

$$\mu^*(x) = \begin{cases} z & \text{if } x \geq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.27}$$

Now consider the following policy for arrivals. At every stage there is a choice between setting the arrival intensity to zero or setting it equal to its max value of y. The value function if we follow this *binary* policy which we have already proved to be optimal in the case when both nature and decision maker is restricted to the class of Markovian policies is:

$$\begin{aligned} V(x) = \max_{\lambda(\cdot)\in\{0,y\}} \mathbb{E}_x \Big[\int_0^\infty & e^{-\delta t} \big(p\mu^*(X_s)\eta^*(X_s) - c\kappa^*(X_s)\lambda(X_s) \\ & + \theta_A \lambda(X_s)(1 - \kappa^*(X_s) + \kappa^*(X_s)\ln\kappa^*(X_s)) \\ & + \theta_D \mu(X_s)(1 - \eta^*(X_s) + \eta^*(X_s)\ln\eta^*(X_s))\big) ds \Big]. \end{aligned} \tag{2.28}$$

$\mu^*(x)$ is as described in (2.27), and $\kappa^*(x), \eta^*(x)$ are as in (2.22). Now suppose we can find a finite constant ν such that $\nu \geq (y\kappa^*(x) + z\eta^*(x)), \forall x$. The existence of such a ν is guaranteed if we look at the expression (2.22) as it is possible to obtain upper and lower bounds on $V(x)$.[1] Given such a ν we can write the following dynamic programming equation (see Bertsekas [Ber95] Ch. 5)

$$\begin{aligned} V(x) = \frac{1}{\delta+\nu} \Big[& p\mu^*(x)\eta^*(x) + \mu^*(x)\theta_D(1 - \eta^*(x) + \eta^*(x)\ln\eta^*(x)) \\ & + (\nu - \mu^*(x))V(x) + \mu^*(x)V(x-1) \\ & + \max\big(- cy\kappa^*(x) + y\theta_A(1 - \kappa^*(x) + \kappa^*(x)\ln\kappa^*(x)) \\ & + y\kappa^*(x)(V(x+1) - V(x)), 0\big). \end{aligned} \tag{2.29}$$

Without loss of generality we can assume that $\delta+\nu = 1$ as it is possible to scale upper bounds z and y appropriately. Substituting the value of $\kappa^*(x)$ and $\eta^*(x)$ from (2.22) in (2.29) and simplifying we obtain:

$$V(x) = \mu^*(x)\theta_D \left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V(x-1))}\right) + \nu V(x) + y\theta_A \max\left(1 - e^{-\frac{1}{\theta_A}(\Delta V(x)-c)}, 0\right). \tag{2.30}$$

To prove the structural properties of $V(x)$ consider the following value-iteration algorithm:

$$V_{n+1}(x) = \mu^*(x)\theta_D \left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V_n(x-1))}\right) + \nu V_n(x) + y\theta_A \max\left(1 - e^{-\frac{1}{\theta_A}(\Delta V_n(x)-c)}, 0\right). \tag{2.31}$$

Such a value-iteration algorithm corresponding to a stochastic game can be shown to converge to the true value function (see [Sha53]). Note that similar iteration equations are

[1] Zero is a lower bound. An upper bound is the value function of the unambiguous problem which can be uniformly upper bounded.

observed in risk sensitive control literature (see [HHM96], [BM02], [CCdO03] and [CM99]).

The set of value-iteration equations can be written more explicitly in the following form:

$$V_{n+1}(x) = z\theta_D \left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V_n(x-1))}\right) + \nu V_n(x) + y\theta_A \max\left(1 - e^{-\frac{1}{\theta_A}(\Delta V_n(x)-c)}, 0\right) \quad (2.32)$$

where by convention $\Delta V_n(-1) = p$ for all n so that $e^{-\frac{1}{\theta_D}(p-\Delta V_n(-1))} = 1$.

Theorem 2. *Suppose we initialize $V_0(x) = 0$ for all x. If we iterate according to equation* (2.32) *then the following holds true for every n*

(a) $\Delta V_n(x) \le p$.

(b) $V_n(x)$ *is increasing in* x, *i.e.,* $\Delta V_n(x) \ge 0$.

(c) $V_n(x)$ *is concave in* x, *i.e.,* $\Delta V_n(x)$ *is decreasing in* x.

Proof. Proof is by induction. By construction the hypothesis holds true for $n = 0$. We now suppose that it holds for $n = k$ and show that it holds for $n = k + 1$.

(a)

$$\begin{aligned}
\Delta V_{k+1}(x) &= z\theta_D \left(\left((1 - e^{-\frac{1}{\theta_D}(p-\Delta V_k(x))}\right) - \left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V_k(x-1))}\right)\right) + \nu(\Delta V_k(x)) \\
&\quad + y\theta_A \left(\max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x+1)-c)}, 0) - \max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0)\right) \\
&\le z\theta_D \left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V_k(x))}\right) + \nu \Delta V_k(x) \\
&\le z(p - \Delta V_k(x)) + \nu \Delta V_k(x) = zp + (\nu - z)p \le zp + (\nu - z)p = p.
\end{aligned}$$

We have used the following facts: $\max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x+1)-c)}, 0) \le \max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0)$ as $\Delta V_k(x)$ is a decreasing function of x, $\left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V_k(x-1))}\right) \ge 0$ as $\Delta V_k(x-1) \le p$ for all x and $1 - e^{-s} \le s$ when $x \ge 0$.

(b)

$$\begin{aligned}
\Delta V_{k+1}(x) &= z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V_k(x-1))} - e^{-\frac{1}{\theta_D}(p-\Delta V_k(x))}\right) + \nu(\Delta V_k(x)) \\
&\quad + y\theta_A \left(\max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x+1)-c)}, 0) - \max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0)\right) \\
&\ge \nu \Delta V_k(x) + y\theta_A \left(\max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x+1)-c)}, 0) - \max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0)\right) \\
&\ge \nu \Delta V_k(x) - \ y\theta_A \max\left(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0\right).
\end{aligned}$$

If $\Delta V_k(x) \le c$, then $\max(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x)-c)}, 0) = 0$ and hence $\Delta V_{k+1}(x) \ge \nu \Delta V_k(x) \ge 0$.

Else if $\Delta V_k(x) \geq c$, then

$$\begin{aligned}\Delta V_{k+1}(x) &\geq \nu \Delta V_k(x) - y\theta_A \left(1 - e^{-\frac{1}{\theta_A}(\Delta V_k(x) - c)}\right) \\ &\geq \nu \Delta V_k(x) - y\theta_A \frac{\Delta V_k(x) - c}{\theta_A} \geq (\nu - y)\Delta V_k(x) \geq 0.\end{aligned}$$

Here we have used the fact that $1 - e^{-s} \leq s$ when $s \geq 0$.

(c)

$$\begin{aligned}\Delta V_{k+1}(x) - \Delta V_{k+1}(x+1) = z\theta_D \left(e^{-\frac{f(x-1)}{\theta_D}} - 2e^{-\frac{f(x)}{\theta_D}} + e^{-\frac{f(x+1)}{\theta_D}}\right) + \nu(\Delta V_k(x) - \Delta V_k(x+1)) \\ + y\theta_A (2\max(1 - e^{-\frac{g(x+1)}{\theta_A}}, 0) - \max(1 - e^{-\frac{g(x)}{\theta_A}}, 0) - \max(1 - e^{-\frac{g(x+2)}{\theta_A}}, 0)),\end{aligned}$$

where $f(x) = p - \Delta V(x)$ and $g(x) = \Delta V(x) - c$. Note that as $0 \leq f(x) \leq p$ and $f(x)$ is increasing in x we have

$$e^{-\frac{f(x-1)}{\theta_D}} - e^{-\frac{f(x)}{\theta_D}} \geq 0.$$

Also as $g(x)$ is decreasing in x,

$$\max(1 - e^{-\frac{g(x+1)}{\theta_A}}, 0) - \max(1 - e^{-\frac{g(x+2)}{\theta_A}}, 0) \geq 0.$$

Therefore

$$\begin{aligned}\Delta V_{k+1}(x) - \Delta V_{k+1}(x+1) &\geq z\theta_D \left(e^{-\frac{f(x+1)}{\theta_D}} - e^{-\frac{f(x)}{\theta_D}}\right) + \nu(\Delta V_k(x) - \Delta V_k(x+1)) \\ &\quad + y\theta_A \left(\max(1 - e^{-\frac{g(x+1)}{\theta_A}}, 0) - \max(1 - e^{-\frac{g(x)}{\theta_A}}, 0)\right) \\ &= z\theta_D \left(e^{-\frac{f(x+1)}{\theta_D}} - e^{-\frac{f(x)}{\theta_D}}\right) + z((p - \Delta V_k(x+1)) - (p - \Delta V_k(x))) \\ &\quad + y\theta_A \left(\max(1 - e^{-\frac{g(x+1)}{\theta_A}}, 0) - \max(1 - e^{-\frac{g(x)}{\theta_A}}, 0)\right) \\ &\quad + y((\Delta V_k(x) - c) - (\Delta V_k(x+1) - c)) \\ &\quad + (\nu - y - z)(\Delta V_k(x) - \Delta V_k(x+1)) \\ &\geq z\theta_D \left(\left(e^{-\frac{f(x+1)}{\theta_D}} + \frac{f(x+1)}{\theta_D}\right) - \left(e^{-\frac{f(x)}{\theta_D}} + \frac{f(x)}{\theta_D}\right)\right) \\ &\quad + y\theta_A \left(\max(1 - e^{-\frac{g(x+1)}{\theta_A}}, 0) - \frac{g(x+1)}{\theta_A}\right) \\ &\quad - y\theta_A \left(\max(1 - e^{-\frac{g(x)}{\theta_A}}, 0) - \frac{g(x)}{\theta_A}\right).\end{aligned}$$

$e^{-s} + s$ is an increasing function of s when $s \geq 0$, so

$$\left(e^{-\frac{f(x+1)}{\theta_D}} + \frac{f(x+1)}{\theta_D}\right) - \left(e^{-\frac{f(x)}{\theta_D}} + \frac{f(x)}{\theta_D}\right) \geq 0.$$

Also $\max(1-e^{-s},0) - s$ is decreasing in s and as $g(x)$ is a decreasing function of x

$$\left(\max(1-e^{-\frac{g(x+1)}{\theta_A}},0) - \frac{g(x+1)}{\theta_A}\right) - \left(\max(1-e^{-\frac{g(x)}{\theta_A}},0) - \frac{g(x)}{\theta_A}\right) \geq 0.$$

Therefore,

$$\Delta V_{k+1}(x) - \Delta V_{k+1}(x+1) \geq 0.$$

□

Hence we have proved here that if we restrict ourselves to the class of Markovian policies and nature is also restricted to choose Markovian policy to hurt the decision maker then a threshold policy is optimum. Specifically we proved that there exists a threshold $b \in [0, \infty]$ such that

$$\begin{cases} c \leq V(x+1) - V(x) \leq p & \text{for } x \leq b \\ V(x+1) - V(x) < c & \text{for } x > b \end{cases} \tag{2.33}$$

Coupled with (2.24) we have the following policy:

$$\lambda^*(x) = \begin{cases} y & \text{if } x \leq b \\ 0 & \text{if } x > b \end{cases} \tag{2.34}$$

Next we will show that the policy remains optimal even if the nature is free to choose any non-Markovian policy. Specifically we prove that if we choose threshold policy and nature is free to choose anything, nature would choose Markovian policy to hurt most.

Theorem 3. *Suppose we choose the input and output intensities according to the equations* (2.34) *and* (2.25)*. Suppose we allow "nature", acting as the adversary, to choose any arbitrary $\mathcal{F}_t$-predictable processes κ_t and η_t to hurt the decision maker so that the expected profit is minimized. Then nature would choose Markovian policy as given by* (2.22)*, where the value function in the equation is the optimal one when both nature and decision maker are allowed to choose only Markovian policies.*

Proof. For any given arbitrary processes $(\kappa_t, \eta_t), t \geq 0$, suppose we consider a situation where nature follows (κ_t, η_t) up to time t and then follows Markovian policy given by (2.22) after that. The value function associated with this (denoted by $\tilde{V}_t$) can be expressed as follows:

$$\tilde{V}_t(x) = \mathbb{E}_x \int_0^t e^{-\delta t}\Big(p\mu^*(X_s)\eta_s - c\lambda^*(X_s)\kappa_s + \theta_A\lambda^*(X_s)(1-\kappa_s+\kappa_s\ln\kappa_s) + \theta_A\mu^*(X_s)(1-\eta_s+\eta_s\ln\eta_s)ds\Big) + \mathbb{E}_x[e^{-\delta t}V(X_t)]. \tag{2.35}$$

To derive the second expectation in above equation consider

$$\int_0^t e^{-\delta s} dV(X_s) = e^{-\delta t}V(X_t) - V(X(0)) + \delta\int_0^t e^{-\delta s}V(X_s)ds. \tag{2.36}$$

Taking expectation on both sides of the equality, we have

$$\mathbb{E}_x \int_0^t e^{-\delta s} dV(X_s) = \mathbb{E}_x[e^{-\delta t}V(X_t)] - V(x) + \mathbb{E}_x\left[\delta\int_0^t e^{-\delta s}V(X_s)ds\right]. \tag{2.37}$$

We can calculate the left most term in the above expression as:

$$\begin{aligned}\mathbb{E}_x \int_0^t e^{\delta s} dV(X_s) &= \mathbb{E}_x \int_0^t e^{-\delta s}\Big([\Delta V(X_s)]dA_s - [\Delta V(X_s-1)]dD_s\Big)\\ &= \mathbb{E}_x \int_0^t e^{-\delta s}\Big(\kappa_s\lambda^*(X_s)\Delta V(X_s) - \eta_s\mu^*(X_s)\Delta V(X_s-1)\Big)ds.\end{aligned} \tag{2.38}$$

The first equality follows from the fact that there are only two possible transitions, upward and downward, and the second equality follows from (2.6). From (2.35) and (2.38) we obtain:

$$\begin{aligned}\mathbb{E}_x[e^{-\delta t}V(X_t)] = V(x) + \mathbb{E}_x \int_0^t e^{-\delta s}\Big(&\kappa_s\lambda^*(X_s)\Delta V(X_s)\\ &- \eta_s\mu^*(X_s)\Delta V(X_s-1) - \delta V(X_s)\Big)ds.\end{aligned} \tag{2.39}$$

From (2.23) we have

$$\begin{aligned}\delta V(X_s) = \theta_A\lambda^*(x)&\Big(1-\exp\Big(-\frac{1}{\theta_A}(\Delta V(X_s)-c)\Big)\Big)\\ &+ \theta_D\mu^*(x)\left(1-\exp\left(-\frac{1}{\theta_D}(p-\Delta V(X_s-1))\right)\right).\end{aligned} \tag{2.40}$$

From (2.39) and (2.40) we obtain

$$\begin{aligned}\mathbb{E}_x[e^{-\delta t}V(X_t)] = V(x) + \mathbb{E}_x \int_0^t e^{-\delta s}\Big(\kappa_s \lambda^*(X_s)\Delta V(X_s) - \eta_s \mu^*(X_s)\Delta V(X_s - 1)\Big)ds \\ - \mathbb{E}_x \int_0^t e^{-\delta s}\theta_A \lambda^*(x)\Big(1 - \exp\Big(-\frac{1}{\theta_A}(\Delta V(X_s) - c)\Big)\Big)ds \\ + \mathbb{E}_x \int_0^t e^{-\delta s}\theta_D \mu^*(x)\left(1 - \exp\left(-\frac{1}{\theta_D}(p - \Delta V(X_s - 1))\right)\right)ds.\end{aligned} \tag{2.41}$$

Substituting $\mathbb{E}_x[e^{-\delta t}V(X(t))]$ from (2.41) to (2.35) we obtain

$$\begin{aligned}\tilde{V}_t(x) = \mathbb{E}_x \int_0^t e^{-\delta s}\Big(\lambda^*(X_s)\Big[-c\kappa_s + \theta_A\Big(\kappa_s(\ln \kappa_s - 1) + e^{-\frac{1}{\theta_A}(\Delta V(X_s) - c)}\Big) + \kappa_s \Delta V(X_s)\Big]\Big)ds \\ + \mathbb{E}_x \int_0^t e^{-\delta s}\Big(\mu^*(X_s)\Big[p\eta_s + \theta_D\Big(\eta_s(\ln \eta_s - 1) + e^{-\frac{1}{\theta_D}(p - \Delta V(X_s - 1))}\Big) \\ - \eta_s \Delta V(X_s - 1)\Big]\Big)ds + V(x).\end{aligned} \tag{2.42}$$

We now prove that the integrands in the expression are non-negative, i.e.,

$$-c\kappa_s + \theta_A\left(\kappa_s(\ln \kappa_s - 1) + e^{-\frac{1}{\theta_A}(\Delta V(X_s) - c)}\right) + \kappa_s \Delta V(X_s) \geq 0 \tag{2.43}$$

and

$$p\eta_s + \theta_D\Big(\eta_s(\ln \eta_s - 1) + e^{-\frac{1}{\theta_D}(p - \Delta V(X_s - 1))}\Big) - \eta_s \Delta V(X_s - 1) \geq 0. \tag{2.44}$$

But this is straightforward as expressions (2.43) and (2.44) are convex in κ and η respectively and from the first order conditions, the values of κ_s and η_s that minimize the integrands are:

$$\begin{aligned}\kappa_s &= e^{-\frac{1}{\theta_A}(\Delta V(X_s) - c)}, \\ \eta_s &= e^{-\frac{1}{\theta_D}(p - \Delta V(X_s - 1))}.\end{aligned} \tag{2.45}$$

Substituting the minimizing value of κ_s in (2.43) and η_s in (2.44) we get zeros. Hence we have proved that

$$\tilde{V}_t(x) \geq V(x). \tag{2.46}$$

□

A similar analysis would prove that if nature chose Markovian policy as defined in (2.22) and we are free to choose any policy, we will again choose threshold policy. So even if we are free to choose anything and nature is restricted to Markovian, we will choose threshold policy. Giving more freedom to nature will only worsen the performance. So it makes sense for us to choose threshold policy. Hence a threshold policy is optimum even if we are free to choose any $\mathcal{F}_t$-predictable intensities and in that case nature would also choose a Markovian policy to hurt us most.

2.3 Effect of Ambiguity Parameter on Threshold Control

In this section we will study the effect of change in ambiguity levels on threshold control. We define the optimal value function explicitly as a function of $\phi := (\theta_A, \theta_D)$ as

$$V^\phi(x) := \max_u \min_{\kappa,\eta} \mathbb{E}_x\Big[\int_0^\infty e^{-\delta t}\big(p\mu_s - c\kappa_s\lambda_s + \theta_A\lambda_s(1-\kappa_s+\kappa_s\ln\kappa_s) + \theta_D\mu_s(1-\eta_s+\eta_s\ln\eta_s)\big)ds\Big]. \tag{2.47}$$

We also define a partial order on ϕ, i.e., $\phi_1 \geq \phi_2$ if $\theta_{1A} \geq \theta_{2A}$ and $\theta_{1D} \geq \theta_{2D}$.

The following property of the value function is obvious from its definition.

Proposition 4. *If $\phi_1 \leq \phi_2$ then $V^{\phi_1}(x) \leq V^{\phi_2}(x)$ for all $x \in \mathcal{N}$.*

Let $b(\phi)$ is the value of optimal threshold control corresponding to the parameter ϕ. We now show that the threshold remains bounded.

Proposition 5. *$b(\phi) < \infty$ for all $\phi \in [0,\infty] \times [0,\infty]$.*

Proof. If $b(\phi) = \infty$ for some ϕ then $\lim_{x\to\infty} V^\phi(x) = \infty$ as $\Delta V^\phi(x) > c$ for all x. But the function $V^\phi(\cdot)$ is uniformly (in x) less than the value function for the unambiguous problem. The value function of the unambiguous problem can be uniformly bounded by setting $\alpha_t = z$ and $\beta_t = 0$ in (2.8). Hence $\lim_{x\to\infty} V^\phi(x) = \infty$ is not possible. □

We can now prove that the optimal threshold control is monotone in θ_A for fixed θ_D.

Proposition 6. *Let $\phi_1 = (\theta_{1A}, \theta_D)$ and $\phi_2 = (\theta_{2A}, \theta_D)$. If $\theta_{1A} < \theta_{2A}$ then $b(\phi_1) \geq b(\phi_2)$.*

Proof. If $x > b(\phi)$ for some $\phi = (\theta_A, \theta_D)$, then $\Delta V^\phi(x) < c$ and hence

$$V^\phi(x) = z\theta_D\left(1 - e^{-\frac{1}{\theta_D}(p-\Delta V^\phi(x-1))}\right) + \nu V^\phi(x) \tag{2.48}$$

which implies

$$\delta V^\phi(x) + z\theta_D\left(e^{-\frac{1}{\theta_D}(p-\Delta V^\phi(x-1))}\right) = z\theta_D. \tag{2.49}$$

Subtracting $\delta V^{\phi}(x-1)$ from both sides we obtain

$$\delta(\Delta V^{\phi}(x-1)) + z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V^{\phi}(x-1))}\right) = z\theta_D - \delta V^{\phi}(x-1). \tag{2.50}$$

Suppose on the contrary $b(\phi_1) < b(\phi_2) < \infty$. By definition of $b(\phi_2)$, $\Delta V^{\phi_2}(b(\phi_2)) \geq c$ but $\Delta V^{\phi_2}(b(\phi_2)+1) < c$. Therefore substituting $x = b(\phi_2)+1$ we obtain the following from eq (2.50)

$$\delta(\Delta V^{\phi_2}(b(\phi_2))) + z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V^{\phi_2}(b(\phi_2)))}\right) = z\theta_D - \delta V^{\phi_2}(b(\phi_2)). \tag{2.51}$$

Also as $b(\phi_1) < b(\phi_2)$

$$\delta(\Delta V^{\phi_1}(b(\phi_2))) + z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V^{\phi_1}(b(\phi_2)))}\right) = z\theta_D - \delta V^{\phi_1}(b(\phi_2)). \tag{2.52}$$

As $\phi_2 \geq \phi_1$, $V^{\phi_2}(b(\phi_2)) \geq V^{\phi_1}(b(\phi_2))$. Hence

$$\delta(\Delta V^{\phi_2}(b(\phi_2))) + z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V^{\phi_2}(b(\phi_2)))}\right) \leq \delta(\Delta V^{\phi_1}(b(\phi_2))) + z\theta_D \left(e^{-\frac{1}{\theta_D}(p-\Delta V^{\phi_1}(b(\phi_2)))}\right). \tag{2.53}$$

The function $\delta s + z\theta_D e^{-\frac{1}{\theta_D}(p-s)}$ is an increasing function of s for $s \geq 0$. So the only way (2.53) can be true is if

$$\Delta V^{\phi_2}(b(\phi_2)) \leq \Delta V^{\phi_1}(b(\phi_2)).$$

As $\Delta V^{\phi_2}(b(\phi_2)) \geq c$, so $\Delta V^{\phi_1}(b(\phi_2)) \geq c$. This contradicts $b(\phi_1) < b(\phi_2)$. □

Numerical experiments also indicate (for various choices of parameters) that threshold value is an increasing function of θ_D. Thus the ambiguity in arrival and ambiguity in departure appear to act in opposite directions (see e.g. figure 2.2). It is therefore important to consider the case when the two ambiguity levels are same. In our numerical experiments for $\theta_A = \theta_D$, the threshold control is increasing in the common ambiguity level (figure 2.3).

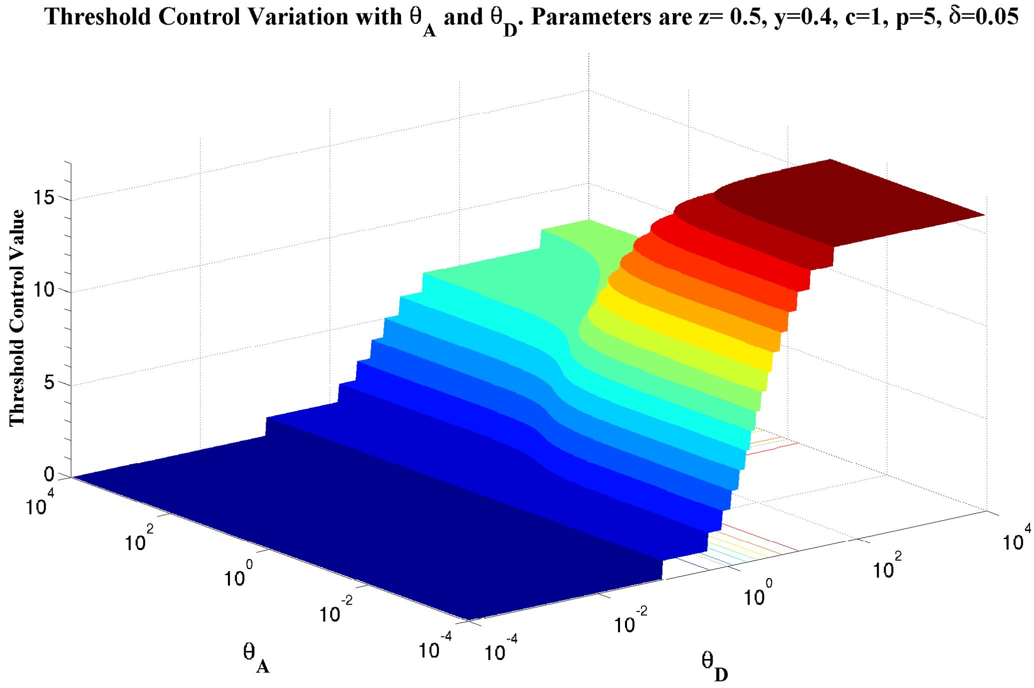

Figure 2.2: Threshold control variation with ambiguity levels.

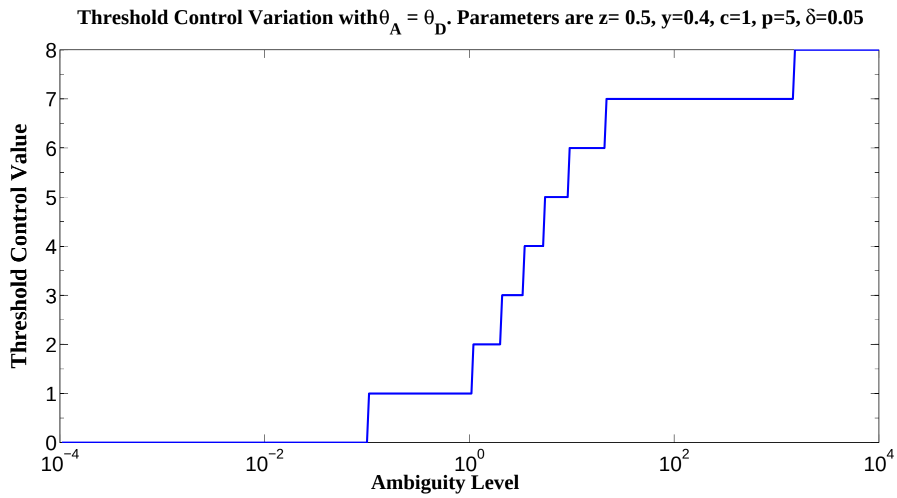

Figure 2.3: Threshold is increasing if $\theta_A = \theta_D$

Chapter 3

Comparison of Different Approaches to Model Uncertainty with Learning

Consider a general discrete time stochastic optimization problem. Let $\boldsymbol{X} := \{X_k\}_{0\leq k\leq n}$ be the underlying stochastic process, where n is the number of decision epochs or planning horizon. The process $\boldsymbol{X}$ is defined on a sample space $(\Omega, \mathcal{F}, \mathcal{F}_k)$ and it subsumes all stochastic processes of interest for the optimization problem. The set Ω is the set of all possible outcomes of $\boldsymbol{X}$ and $\mathcal{F}$ is a sigma algebra associated with Ω. The set $\mathcal{F}_0$ (sigma algebra) can be understood as the set of all information available at time 0, such as past demand data or a priori subjective belief of an expert about the future demand. The set $\mathcal{F}_k$ contains all possible information at time k. It is important to note that the information available to decision maker at any time k may be a strict subset of $\mathcal{F}_k$. For example, the decision maker may know only the past sales data and not the actual demand data. Therefore, we make a distinction and denote by $\mathcal{I}_k$ the set of information available to the decision maker at time k.

Let y_k be the decision made at the beginning of time k. The decision is made based on the knowledge of the information set $\mathcal{I}_{k-1}$. Let $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ be the policy of the decision maker and $\mathcal{Y}$ be the set of all admissible policies. We consider the following generic optimization problem:

$$\max_{\boldsymbol{y}\in\mathcal{Y}} \mathbb{E}\left[\psi\left(\boldsymbol{y}, \boldsymbol{X}\right)\right)\left|\mathcal{I}_0\right] \tag{3.1}$$

In solving a stochastic optimization problem such as (3.1), a four steps procedure is usually followed:

1. *Choose the best objective to maximize*: Depending on the problem and risk preference a function ψ of the policy $\boldsymbol{y}$ and stochastic process $\boldsymbol{X}$ is chosen. This function is maximized in some way (for example, expected value) given the initial information set $\mathcal{I}_0$. Henceforth we drop $\mathcal{I}_0$ for convenience of presentation. Unless otherwise mentioned

it should be understood that the objective is maximized given the initial information $\mathcal{I}_0$.

2. *Make suitable assumptions about the underlying stochastic process*: Let $\mathcal{P}$ be the probability measure that governs the stochastic process $\boldsymbol{X}$. To estimate $\mathcal{P}$, some statistical assumptions about the stochastic process are made based on previous knowledge, expert opinions or mathematical convenience. For example, we may assume that the demand in an inventory control problem is independent and identically distributed with exponential distribution.

3. *Estimate parameters/distributions based on assumption*: Given the assumptions about the stochastic process, the parameters of the model are estimated or an estimate $\mathcal{P}_0$ of the true probability measure $\mathcal{P}$ is calculated using past data. For instance, in the inventory example with i.i.d. and exponential distributed demand, the mean of past demand data is an estimate of the parameter of the exponential distribution. One may also use subjective *Bayesian priors* where the estimates are considered random and have a probability distribution in contrast to point estimates in classical statistics.

4. *Solve the problem using estimates*: The problem is then solved assuming that the estimate $\mathcal{P}_0$ is the true probability measure. It is hoped that the solution obtained using $\mathcal{P}_0$ would be close to the true optimal (when $\mathcal{P}$ is known) in some sense.

One subjective element in the above procedure is the choice of statistical assumptions. The assumptions are made in hope that the resulting analytical model is close to the actual system. But in many cases this may not be the case, because even if the assumptions are correct, one may not get a good estimate of the model parameters if the data is limited or the past data is not a true reflection of future. In subsequent sections we discuss the effects of incorrect assumptions or errors in model. We discuss various ways to model these errors and present a comparison of these approaches.

Throughout the paper we present the newsvendor problem to discuss and compare different ideas. The simplicity of the problem allows us to concentrate on ideas without worrying about calculations and numerical tractability. The formal statement of the newsvendor problem is as follows.

Newsvendor Problem

Consider a perishable item which is purchased at a cost of c per unit and sold at a price of s per unit. The demand for the product is random and can take value anywhere in $[0, \infty)$. The salvage value of the unsold item at the end of the period is 0 and thus no inventory is carried over. Suppose the order quantities in the first n periods are $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ and the demands are $\boldsymbol{D} = \{D_1, D_2, \ldots D_n\}$. Then the profit $\psi(\boldsymbol{y}, \boldsymbol{D})$ is

$$\psi(\boldsymbol{y}, \boldsymbol{D}) = \sum_{i=1}^{n} \{s \min\{y_i, D_i\} - cy_i\}. \tag{3.2}$$

We also assume that the decision maker is risk neutral and wants to maximize the expected profit in each period. Thus the problem is to maximize

$$\phi(\boldsymbol{y}) = \mathbb{E}\psi(\boldsymbol{y}, \boldsymbol{D}) = \sum_{i=1}^{n} \mathbb{E}\left[\{s \min\{y_i, D_i\} - cy_i\}\right]. \tag{3.3}$$

Throughout the chapter, while comparing various approaches, we make a statistical assumption that $D_1, D_2, \ldots, D_n$ are i.i.d with a continuous distribution F_D. This commonly made assumption may not always be valid.

In addition, when required we make an assumption that the demand is exponentially distributed with mean θ, i.e. $F_D(x) = 1 - \exp(-\frac{x}{\theta}), D \geq 0$.

The rest of the chapter is organized as follows: In Section 3.1, we review some of the classical modeling approaches in operations management.[1] In Section 3.2, effects of wrong model on the modeling methodologies mentioned in Section 3.1 are discussed. In Section 3.3, we review some common uncertainty sets based on data, which are used to describe a collection of models. We compare the performance of different robust optimization models with various uncertainty sets in Section 3.4.

3.1 Classical Modeling

3.1.1 Deterministic Modeling

Many of the earlier models in operations management were fully deterministic. Mathematically, deterministic modeling is equivalent to choosing a specific instance $\omega_0 \in \Omega$ and maximizing:

$$\max_{\boldsymbol{y} \in \mathcal{Y}} \psi(\boldsymbol{y}, \boldsymbol{X}(\omega_0)) \tag{3.4}$$

A practical implementation of this policy would require a method to choose ω_0 based on initial information. Typical forecasting methods like moving average or exponential smoothing can be used to choose ω_0. For example, the initial information in period may consist of past values of the stochastic process $\boldsymbol{X}$ i.e., $X_0, X_{-1}, X_{-2}, \ldots, X_{-m}$. One possible way to

[1] This chapter is partly based on IEOR 290A course lecture notes taken by the author in Spring 2006 at University of California, Berkeley

utilize this information in a deterministic way is:

$$X_k(\omega_0) = \sum_{i=-m}^{0} X_i, \qquad k = 1, 2, \ldots, n. \tag{3.5}$$

A large number of works in mathematical programming and robust optimization use deterministic modeling. The resulting problem often falls within well defined concepts of mathematical programming, and therefore can be solved efficiently using standard tools and software. However, unless the variability in the stochastic process is low, a deterministic approach is not likely to result in a good solution.

Suppose the demand in the newsvendor problem is assumed to be deterministic, i.e., it is $D_k(\omega_0)$. Then, the optimal order quantity y_k is $D_k(\omega_0)$. Hence the newsvendor profit $\psi^D(\omega_0)$ in deterministic case is:

$$\psi^D(\omega_0) = \sum_{k=1}^{n} \left(s \min\{D_k, D_k(\omega_0)\} - cD_k(\omega_0)\right) \tag{3.6}$$

One possible way to get the forecast $D_k(\omega_0)$ is to use the average of data from past l_k periods. Then,

$$D_k(\omega_0) := \bar{D}_k^{l_k} := \frac{1}{l_k} \sum_{i=k-l_k}^{k-1} D_i \tag{3.7}$$

Using the forecast, the profit is:

$$\psi^D = \sum_{k=1}^{n} \left(s \min\left\{D_k, \bar{D}_k^{l_k}\right\} - c\bar{D}_k^{l_k}\right) \tag{3.8}$$

3.1.2 Stochastic Modeling

A typical stochastic model in operations management assumes a probability measure P on $(\Omega, \mathcal{F}, (\mathcal{F}_k))$. Often some statistical assumptions such as *i.i.d.* are made about the stochastic process. It is desirable to specify the stochastic process under minimal assumptions. However, a weaker assumption usually means worse model calibration due to lack of past data or subjective information. The generic optimization problem given the probability measure P is:

$$\max_{\boldsymbol{y} \in \mathcal{Y}} \mathbb{E}_P \left[\psi(\boldsymbol{y}, \boldsymbol{X})\right] \tag{3.9}$$

The statistical assumptions and modeling can further be classified into parametric and non-parametric modeling. Each of these can further be subdivided into *Bayesian* or *Fre-*

quentist approach.

Parametric approach

The probability distribution P_θ belongs to a set $\mathcal{P}_\Theta$, characterized by a finite dimensional parameter θ. A frequentist approach assumes that the parameter θ is fixed but unknown. Let $\boldsymbol{y}(\theta)$ be the solution of the following optimization problem:

$$\max_{\boldsymbol{y}\in\mathcal{Y}} \mathbb{E}_{P_\theta}\left[\psi(\boldsymbol{y}, \boldsymbol{X})\right] \tag{3.10}$$

To implement a policy using a frequentist approach, one finds an estimate of θ based on initial information. The estimate $\hat{\theta}\,(\mathcal{I}_0)$ of θ is typically estimated by statistical techniques such as using maximum likelihood estimator or uniformly minimum variance unbiased (UMVU) estimator. In this case the implemented solution $\boldsymbol{y}^{PF}$ is:

$$\boldsymbol{y}^{PF} = \boldsymbol{y}(\hat{\theta}\,(\mathcal{I}_0)). \tag{3.11}$$

As a specific example of parametric modeling, we consider the newsvendor problem with *i.i.d.* and exponentially distributed demand with mean θ. In this case, the objective function and the optimal policy (see [LS05]) are:

$$\mathbb{E}_{P_\theta}\left[\psi(\boldsymbol{y}, \boldsymbol{D})\right] = \sum_{k=1}^{n}\left(s\theta\left(1 - \exp\left\{-\frac{y_k}{\theta}\right\}\right) - cy_k\right), \tag{3.12}$$

and

$$y_k^{PF}(\theta) = \theta \ln\left(\frac{s}{c}\right). \tag{3.13}$$

For an exponential distribution, the sample mean is the UMVU estimator of θ. Hence one can use the sample mean of the observed data to estimate θ. The *implemented* ordering policy is then

$$\hat{y}_k^{PF} = \bar{D}_k^{l_k} \log\left(\frac{s}{c}\right), \quad k = 1, 2, \ldots, n, \tag{3.14}$$

with profit

$$\psi^{PF} = \sum_{k=1}^{n}\left\{s \min\left\{D_k, \bar{D}_k^{l_k} \log\left(\frac{s}{c}\right)\right\} - c\bar{D}_k^{l_k} \log\left(\frac{s}{c}\right)\right\}, \tag{3.15}$$

where, $\bar{D}_k^{l_k} = \frac{1}{l_k}\sum_{i=k-l_k}^{k-1} D_i$.

In parametric *Bayesian*, one assumes the parameter θ of the underlying distribution to be random and one chooses an a priori distribution for the parameter θ. Suppose the a priori distribution for θ is $F(\theta)$, $\theta \in \Theta$. The objective function in this case is

$$E_\Theta\left[\psi\left(\boldsymbol{y}, \boldsymbol{X}(\theta)\right)|\mathcal{I}_0\right] = \int_{\theta\in\Theta} \psi\left(\boldsymbol{y}, \boldsymbol{X}(\theta)\right) dF(\theta). \tag{3.16}$$

Let

$$\boldsymbol{y}^{PB}(\mathcal{I}_0) = \arg\max_{\boldsymbol{y}\in\mathcal{Y}}\ E_\Theta\left[\psi\left(\boldsymbol{y}, \boldsymbol{X}(\theta)\right)|\mathcal{I}_0\right] \tag{3.17}$$

When the re-optimization is done at every step, as in newsvendor problem, a popular choice of subjective prior is the conjugate of the demand distribution (e.g. [Azo85]). When the demand is exponentially distributed, the conjugate choice of prior is gamma distribution. The probability density of gamma distribution with parameters α and β (which are subjectively chosen) and the rate $\frac{1}{\theta}$ is:

$$f(\theta) = \frac{(\frac{\beta}{\theta})^{\alpha+1}}{\beta\Gamma(\alpha)} \exp\left\{-\frac{\beta}{\theta}\right\}, \theta \geq 0. \tag{3.18}$$

Suppose at every time instant k the information we have or want to use is the last l_k periods of past data. Straightforward algebra will reveal that

$$\hat{y}_k^{PB} = \left(\beta + l_k\bar{D}_k^{l_k}\right)\left(\left(\frac{s}{c}\right)^{\frac{1}{\alpha+l_k}} - 1\right), \tag{3.19}$$

with profit

$$\hat{\psi}^{PB} = \sum_{k=1}^{n}\left\{s\min\left\{D_k, \left(\beta + l_k\bar{D}_k^{l_k}\right)\left(\left(\frac{s}{c}\right)^{\frac{1}{\alpha+l_k}} - 1\right)\right\} - c\left(\beta + l_k\bar{D}_k^{l_k}\right)\left(\left(\frac{s}{c}\right)^{\frac{1}{\alpha+l_k}} - 1\right)\right\}. \tag{3.20}$$

Non-Parametric Approach

Non-parametric modeling do not describe the probability distribution of stochastic process using finite number of parameters, and hence make less statistical assumptions. Let $\hat{P}$ be an estimate (which is calculated non-parametrically) of distribution of $\boldsymbol{X}$, then the optimal non-parametric solution $y^{NP}(\hat{P})$ is:

$$\boldsymbol{y}^{NP}(\hat{P}) = \arg\max_{y\in\mathcal{Y}} \mathbb{E}_{\hat{P}}\left[\psi(\boldsymbol{y}, \boldsymbol{X})\right] \tag{3.21}$$

To implement the non-parametric solution we need a method to calculate the distribution $\hat{P}$. Empirical distribution of past data and kernel density estimation (for continuous distributions) are two of the methods used.

For the newsvendor problem observe that the optimal order quantity $y_k^{NP}(F_{D_k})$ for demand distribution F_{D_k} is given by

$$y_k^{NP}\left(F_{D_k}\right) = \bar{F}_{D_k}^{inv}\left(\frac{c}{s}\right), \tag{3.22}$$

where $\bar{F}_{D_k}^{inv}$ is the inverse of the survival function ($\bar{F}_{D_k} = 1 - F_{D_k}$) of the demand. Suppose in period k we use the empirical distribution based on last l_k demand data points. Let $D_{[0]}^k = 0$ and $D_{[r]}^k$ be the r-th order statistic of $\{D_{k-1}, D_{k-2} \ldots, D_{k-l_k}\}$, $r = 1, 2, \ldots, l_k$. Since the demand is assumed to be continuous, we set

$$\hat{\bar{F}}_{D_k}(x) = 1 - \frac{1}{l_k}\left\{r - 1 + \frac{x - D_{[r-1]}^k}{D_{[r]}^k - D_{[r-1]}^k}\right\},\ D_{[r-1]}^k < x \le D_{[r]}^k,\ r = 1, 2, \ldots, l_k. \tag{3.23}$$

Then the implemented order quantity $\hat{\pi}^g$ based on the empirical distribution is:

$$\hat{y}_k^{NP} = \hat{\bar{F}}_{D_k}^{inv}\left(\frac{c}{s}\right) = D_{[\hat{r}-1]}^k + \hat{a}(D_{[\hat{r}]}^k - D_{[\hat{r}-1]}^k), \tag{3.24}$$

where $\hat{r} \in \{1, 2, \ldots, l_k\}$ satisfies

$$l_k\left(1 - \frac{c}{s}\right) < \hat{r} \le l_k\left(1 - \frac{c}{s}\right) + 1, \tag{3.25}$$

and

$$\hat{a} = l_k\left(1 - \frac{c}{s}\right) + 1 - \hat{r}. \tag{3.26}$$

3.2 Modeling Errors

In this section we discuss the effect of modeling errors on the models mentioned in Section 3.1. We distinguish between two types of errors: (i) calibration error and (ii) error due to wrong statistical assumptions. Even if our statistical assumptions about the nature of stochastic process is right, we may still have an error in model due to limited amount of data available for calibration.

3.2.1 Calibration Error

Let ϕ_k denotes the expected profit in period k, specifically ϕ_k^D denotes the expected profit in period k using deterministic policy; ϕ_k^{PF} represents the expected profit in period k using parametric frequentist method and so on. Mathematically:

$$\phi_k^{PF} = \mathbb{E}_P\left[s \min\left\{D_k, \bar{D}_k^{l_k} \log\left(\frac{s}{c}\right)\right\} - c\bar{D}_k^{l_k} \log\left(\frac{s}{c}\right)\right] \tag{3.27}$$

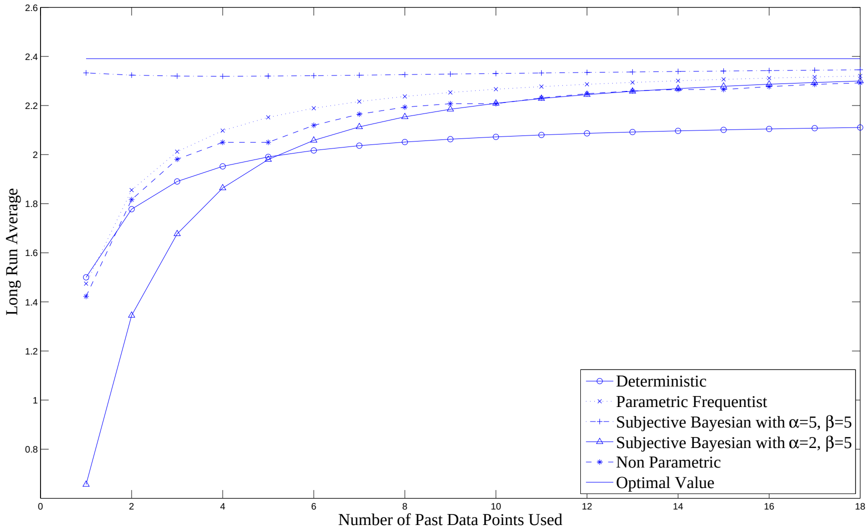

Figure 3.1: Comparison of classical modeling methods discussed in Section 3.1. Values of parameters are s=5, c=1, $\theta = 1$.

The probability measure P is the true probability measure of the stochastic process $\boldsymbol{D}$. Let the true probability measure is i.i.d. and exponentially distributed as assumed in the parametric model. The optimal profit as a function of mean demand parameter θ is given by:

$$\phi(\theta) = (s-c)\theta - c\theta \log \frac{s}{c} \tag{3.28}$$

If we believe the assumption that the demand is i.i.d., then it make sense to use $l_k = k$. It can be shown for $l_k = k$:

$$\phi_k^D = s\theta\left(1 - \left(\frac{k}{k+1}\right)^n\right) - c\theta \tag{3.29}$$

$$\phi_k^{PF} = s\theta\left(1 - \left(\frac{k}{k+\ln(s/c)}\right)^n\right) - c\theta \tag{3.30}$$

$$\tag{3.31}$$

Figure 3.1 shows the performance of various models discussed in Section 3.1. All the models except for the deterministic model converge to the optimal solution asymptotically. A Bayesian model with good subjective prior ($\alpha = 5, \beta = 5$) has very good small sample performance whereas if a Bayesian prior with $\alpha = 2, \beta = 5$ is used the performance is even worse than non-parametric model. Hence the choice of subjective prior is very relevant to the performance of a Bayesian model.

The other major issue in operations management is the non stationarity of stochastic process. In presence of non-stationarity we can only trust past few data points for calibration. Suppose $l_k = m$, where m a constant independent of k. Let $\hat{\phi}(m)$ is the long run average of the profit in a period. Specifically:

$$\hat{\phi}^D(m) = \lim_{n \leftarrow \infty} \frac{1}{n} \sum_{k=1}^{n} \left[s \min \left\{ D_k, \bar{D}_k^m \log\left(\frac{s}{c}\right) \right\} - c\bar{D}_k^m \log\left(\frac{s}{c}\right) \right]. \tag{3.32}$$

$\hat{\phi}^{PF}$, $\hat{\phi}^{PB}$ and $\hat{\phi}^{NP}$ are defined similarly. The following can be shown (see [LS05]) if the demand process is exponentially distributed:

$$\hat{\phi}^D(m) = (s-c)\theta - s\theta \left(\frac{m}{m+1}\right)^m, \tag{3.33}$$

$$\hat{\phi}^{PF}(m) = s\theta \left(1 - \left(\frac{m}{m + \ln(\frac{s}{c})}\right)^m\right) - c\theta \ln\left(\frac{s}{c}\right), \tag{3.34}$$

$$\hat{\phi}^{PB}(m) = s\theta \left(1 - \left(\frac{\theta}{\left(\frac{s}{c}\right)^{\frac{1}{\alpha+m}} + \theta - 1}\right)^m \exp\left\{-\frac{\beta}{\theta}\left(\left(\frac{s}{c}\right)^{\frac{1}{\alpha+m}} - 1\right)\right\}\right) \tag{3.35}$$

$$- c(\beta + m\theta)\left(\left(\frac{s}{c}\right)^{\frac{1}{\alpha+m}} - 1\right), \tag{3.36}$$

$$\hat{\phi}^{NP} = c\theta \left\{ \frac{s}{c}\left(1 - \left(\frac{m - \hat{r} + 2}{m+1}\right)\left(\frac{m - \hat{r} + 1}{m - \hat{r} + 1 + \hat{a}}\right)\right) - \sum_{k=1}^{\hat{r}-1} \frac{1}{m-k+1} - \frac{\hat{a}}{m - \hat{r} + 1} \right\}. \tag{3.37}$$

3.2.2 Error due to Wrong Statistical Assumptions

Performance of a model may suffer not only due to lack of sufficient data to calibrate the model, but also due to wrong model assumptions. In all models of Section 3.1, we have assumed that the demand process is *i.i.d.* and exponentially distributed. In this section we consider the effect on various models if each of the above assumptions are not true. In particular, we consider two different ways a model is misspecified:

(i) the demand is gamma distributed with scale parameter 0.1 and shape parameter 10 instead of being exponentially distributed. Figure 3.2 shows the performance of various models discussed in Section 3.1. All the models except for the non-parametric model converge to the wrong order quantity due to model misspecification.

(ii) the observed demand is a function of order quantity as in the case of a censored

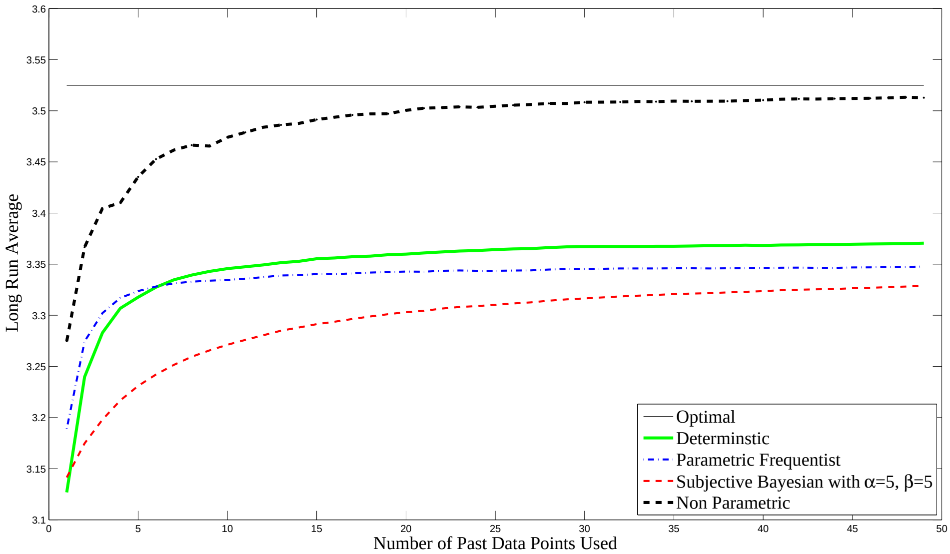

Figure 3.2: Effect of wrong model on long run profit. Demand is gamma distributed with scale parameter 0.1 and shape parameter 10 instead of exponential.

newsvendor problem. In a censored newsvendor problem sales data is observed instead of actual demand data. We consider the effect of using sales data instead of actual demand data to construct estimate of demand distribution. Figure 3.3 shows the performance of all the models using censored data. In all the models, using censored data leads to progressive worsening of demand distribution, which in turn leads to smaller (stochastically) order quantity. Eventually in all the models the order quantities, as well as the estimate of demand distribution, converge to zero.

3.3 Flexible Modeling using Variable Uncertainty set

As discussed in Section 3.2, it is clear that classical models of operations management performs poorly due to error in calibration and also due to model misspecification. Therefore, there is a need to use models that explicitly take into account the errors due to calibration and misspecification. One possible way to do this is to consider more than one model while making decisions. In our optimization problem it is equivalent to specifying multiple probability measure for the stochastic process $\boldsymbol{X}$. So instead of specifying a single probability measure P, we specify a collection of probability measure $\mathcal{P}$ and assume that the true probability measure lies in the set $\mathcal{P}$ or there is a probability measure in the set which is

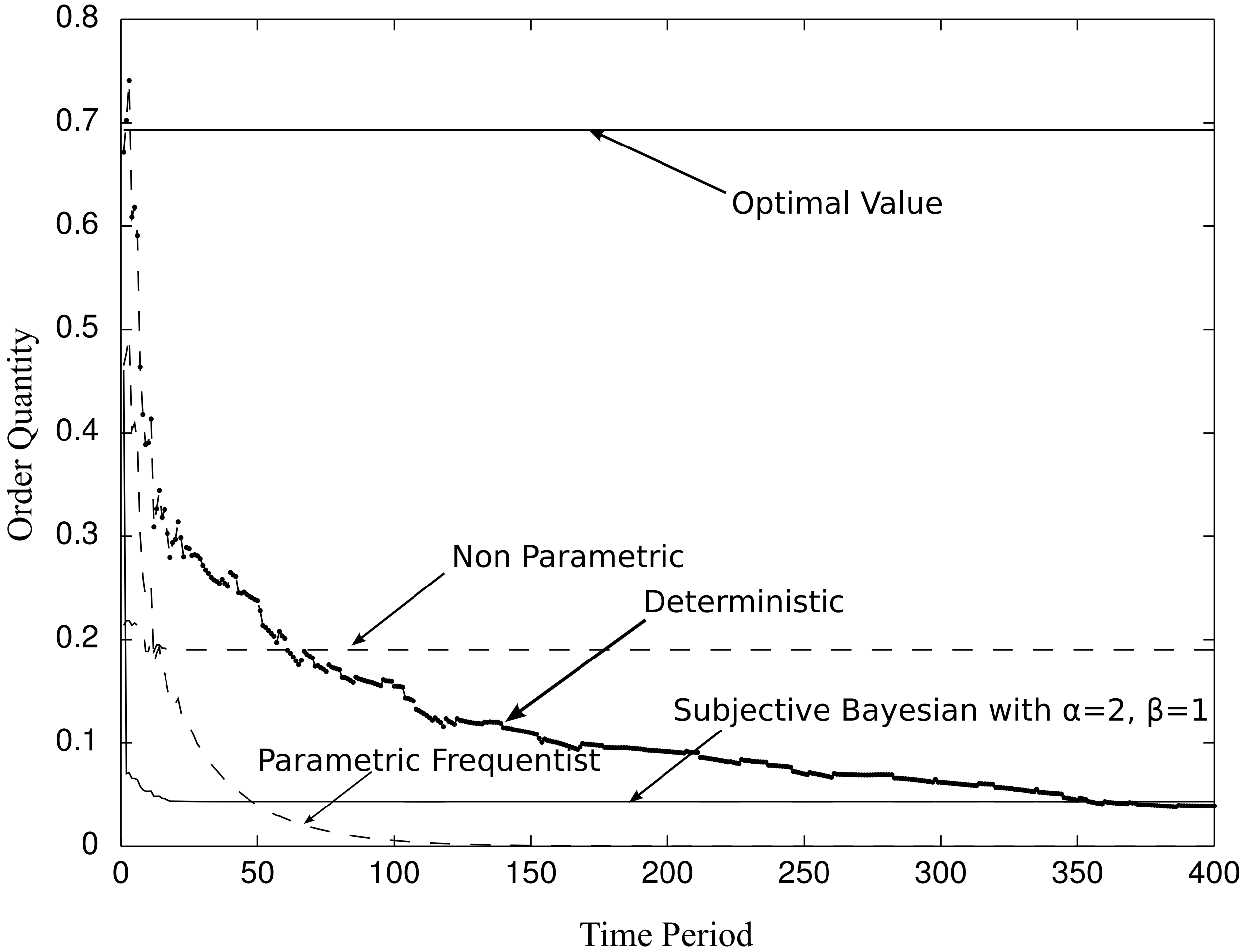

Figure 3.3: Spiral down effect observed using censored demands

very close (in some sense) to the true probability measure. The variable uncertainty set $\mathcal{P}$ can be defined in both parametric and non-parametric ways.

3.3.1 Deterministic Variable Uncertainty Set

The deterministic uncertainty set is a collection of degenerate probability measure, i.e.,

$$\mathcal{P}^D = \{P_\omega^D,\, \omega \in \Omega\}. \tag{3.38}$$

This is essentially equivalent to defining a set of values that $\boldsymbol{X}$ can take. Let the set of possible values be $\mathcal{X}$. Then one considers a collection of problems:

$$\psi(\boldsymbol{y}, \boldsymbol{X}),\, X \in \mathcal{X}. \tag{3.39}$$

In the newsvendor problem, where $\bar{D}_k^{l_k}$ is the forecast of the future, the deterministic uncertainty set is given by:

$$\mathcal{D}_k^D = \{D : \bar{D}_k^{l_k} - a_k \leq D \leq \bar{D}_k^{l_k} + b_k\} \tag{3.40}$$

In practice it is hard to come up with values of a_k and b_k unless there are natural limits or engineering application. A natural limit can be found by using population size, but using the population size is not expected to give a good solution as we shall see in the next section.

Another possible way to get a_k's and b_k's is to use confidence interval associated with $\bar{D}_k^{l_k}$ assuming l_k is sufficiently large so that distribution of $\bar{D}_k^{l_k}$ is close to normal. The values of a_k and b_k then would be:

$$a_k = \bar{D}_k^{l_k} - z_{\alpha/2}\frac{S_k^{l_k}}{l_k}, b_k = \bar{D}_k^{l_k} + z_{\alpha/2}\frac{S_k^{l_k}}{l_k}, \tag{3.41}$$

for appropriate value of α, $0 < \alpha < 1$. $S_k^{l_k}$ is the standard deviation of last l_k entries.

3.3.2 Parametric Variable Uncertainty Set

A possible way to consider the error in calibration of a parametric model is to consider a family of probability measures based on different values of parameter. In a parametric frequentist setting the collection of measures we consider could be

$$\mathcal{P}^{PF} = \{P_\theta^{PF},\, \theta \in \Theta\}, \tag{3.42}$$

for some set Θ of parameter values.

It is common in robust optimization to specify set Θ using a confidence set for θ. Let $0 < \alpha < 1$ and $t\,(\mathcal{I}_0)$ be an estimator of θ given initial information $\mathcal{I}_0$. The uncertainty set

Θ is chosen so that

$$P\left(t\left(\mathcal{I}_0\right) \in \Theta\right) = 1 - \alpha. \tag{3.43}$$

One problem here is how to choose an appropriate value of α, as the solution is dependent on α. A robust solution is desirable if it is not too sensitive to the choice of α.

For the newsvendor problem if the past l_k periods data is assumed to be *i.i.d.* exponential then it can be shown that:

$$\frac{2l_k}{\theta}\bar{D}_k^{l_k} =^d \chi^2_{2l_k}, \tag{3.44}$$

where $\chi^2_{2l_k}$ is a chi-square random variable with $2l_k$ degrees of freedom. From (3.43) and (3.44) it is immediate that

$$P\left(\frac{2l_k\bar{D}_k^{l_k}}{\chi^2_{2l_k,\frac{\alpha}{2}}} \leq \theta \leq \frac{2l_k\bar{D}_k^{l_k}}{\chi^2_{2l_k,1-\frac{\alpha}{2}}}\right) = 1 - \alpha, \tag{3.45}$$

where

$$P\{\chi^2_{2l_k,\beta} \leq \chi^2_{2l_k}\} = \beta, \beta \geq 0.$$

A $(1-\alpha)$ 100 % confidence interval for θ is $\left(\frac{2l_k\bar{D}_k^{l_k}}{\chi^2_{2l_k,\frac{\alpha}{2}}}, \frac{2l_k\bar{D}_k^{l_k}}{\chi^2_{2l_k,1-\frac{\alpha}{2}}}\right)$ which can be used as set Θ.

A Bayesian way to specify the variable uncertainty set is to consider hierarchical models. If the parameters of prior distribution are not well specified then in true Bayesian sense we should consider the parameters of prior distribution itself to be random and specify a probability measure on parameters. We can consider multiple levels of hierarchy, though in practice it is not advantageous to have more than two or three levels of hierarchy.

3.3.3 Nonparametric Variable Uncertainty Set

There are multiple ways to specify a non-parametric uncertainty set. One particular way is to specify some properties about the distribution of the stochastic process and consider the class of all probability measures that satisfy those properties. For example a non-parametric variable uncertainty set can be a set of all probability measures such that the mean vector and covariance matrix of the stochastic process are some particular values, i.e.,

$$\mathcal{P} = \{P : \mathbb{E}_P\left[\boldsymbol{X}\right] = \boldsymbol{\mu}, \mathbb{E}_P\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right] = Q.\} \tag{3.46}$$

A non-parametric variable uncertainty set can also be specified by choosing a ball of probability measure centered around a nominal probability measure. All measures in the ball are within some "distance" from the nominal probability measure. Therefore, to completely specify a non-parametric variable uncertainty set we need a nominal probability measure and a measure of distance between probability measures. One way to specify nominal probability measure is to choose it non-parametrically as in Section 3.1.2, although we can specify a parametric nominal probability measure while still choosing the variable uncertainty set non-parametrically. If $\hat{P}$ is the nominal probability measure and $d(P, \hat{P})$ is a measure of distance between P and $\hat{P}$, then a non-parametric variable uncertainty set is:

$$\mathcal{P} = \{P : d(P, \hat{P}) \leq \alpha\}, \tag{3.47}$$

for some parameter value α. Ideally we would like to choose α so that the true probability measure is in $\mathcal{P}$. In practice it may be hard to choose such an α but we hope that our solution is not too sensitive to α and a reasonable choice of α would give us a robust enough solution.

Some of the distance measures commonly used are listed below:

Kullback-Leibler(KL) Divergence or Relative Entropy

$$d_{KL}(P, \hat{P}) = \int_{\Omega} \log \left(\frac{dP}{d\hat{P}} \right) dP, \tag{3.48}$$

where $\frac{dP}{d\hat{P}}$ is so called Radon-Nikodym derivative. Kullback-Leibler divergence is a popular measure to specify variable uncertainty sets specifically in dynamic models. It has a useful property that it is sum separable for product measures. It is convex in P and takes values in $[0, \infty]$. However, it is not a metric (it is not symmetric in $(P, \hat{P})$ and does not satisfy the triangle inequality). Another caveat is if there exists a set A, however small, such that $P(A) > 0$ but $P(A) = 0$, then $d_{KL}(P, \hat{P}) = \infty$.

Total Variation Distance

Total variation distance is defined as:

$$d_{TV}(P, \hat{P}) = \sup\{|P(A) - \hat{P}(A)| : A \in \mathcal{F}\} \tag{3.49}$$

Total variation distance defines a norm on the space of probability measures.

Lévy (Prokhorov) Metric

Lévy metric, which is a special case of Lévy-Prokhorov metric, between two probability distributions F and $\hat{F}$ is defined as

$$d_L(F, \hat{F}) = \inf\{h : F(x-h) - h \leq \hat{F}(x) \leq F(x+h) + h; h > 0; x \in \mathcal{R}\}. \tag{3.50}$$

3.4 Robust Optimization Models

Assuming that the true probability measure can be any one from the variable uncertainty set, robust optimization methods use suitable objective functions to derive solutions which are hoped to be robust enough to model misspecification and calibration error. Some of the most commonly used objective functions are max-min, max-min regret and max-min competitive ratio.

3.4.1 Max-min Objective

The max-min or worst case objective is the most commonly used robust objective in operations management literature. The optimization problem is:

$$\max_{\boldsymbol{y} \in \mathcal{Y}} \min_{P \in \mathcal{P}} \mathbb{E}\left[\psi(\boldsymbol{y}, \boldsymbol{X})\right]. \tag{3.51}$$

This approach finds the best solution assuming the true model is the worst one possible. If the variable uncertainty set is large then the solution might be too conservative. On the other hand if the uncertainty set is too small then the solution may not be robust at all. In practice we want a solution that is not too sensitive to the variable uncertainty set; min-max as a robust objective fails to achieve that as we shall see in the newsvendor problem context.

Max-min with Deterministic Uncertainty Set

If the demand in period k can take values in $[a_k, b_k]$, then the robust optimization problem is:

$$\max_{\boldsymbol{y} \geq 0} \min_{a_k \leq D_k \leq b_k} \sum_{k=1}^{n} \left(s \min\{D_k, y_k\} - c y_k\right). \tag{3.52}$$

Since the inner minimization is monotone in D_k, the optimization problem (3.52) can be written as:

$$\max_{\boldsymbol{y} \geq 0} \sum_{k=1}^{n} \left(s \min\{a, y_k\} - c y_k\right), \tag{3.53}$$

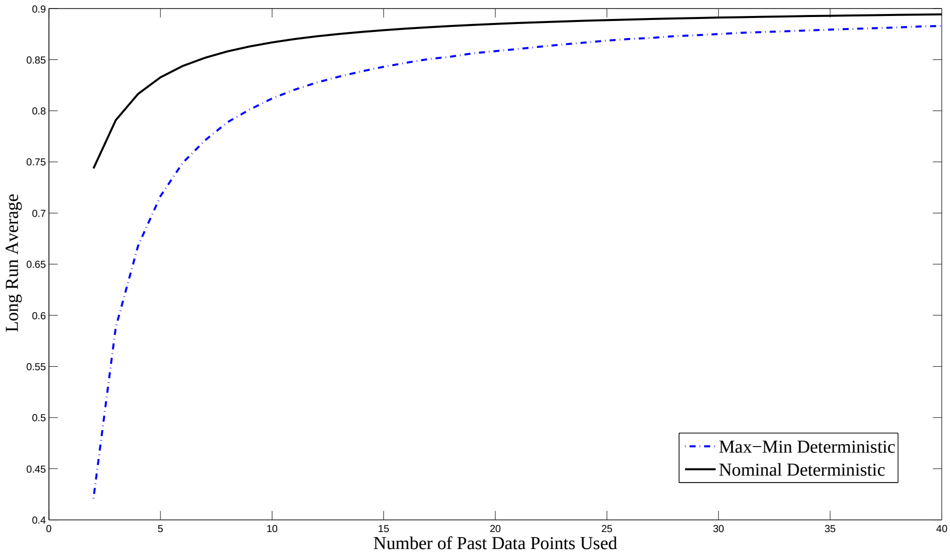

Figure 3.4: Performance of Max-min robust optimization objective with deterministic uncertainty set.

with optimum solution

$$y_k^{R:D} = a_k, k = 1, 2, \ldots, m. \tag{3.54}$$

To implement the solution we need values for a_k's. A natural boundary is $a_k = 0$ but this would give a profit of 0, which is clearly a pessimistic solution.

Instead if we use the value of a_k as defined in (3.41), the profit $\psi^{R:D}$ is:

$$\psi^{R:D} = \sum_{k=1}^{n} s \min\left\{ \left[\bar{D}_k^{l_k} - z_{\alpha/2} S_k^{l_k}\right]^+, D_k \right\} - c\left[\bar{D}_k^{l_k} - z_{\alpha/2} S_k^{l_k}\right]^+. \tag{3.55}$$

As we see from Figure 3.4, the performance of the deterministic max-min optimization is even worse than deterministic optimization for 90% confidence interval estimates for α. We can, of course, improve the performance by choosing different values of confidence level. However, first it is difficult to say which value of confidence level is good a priori. In addition, as we see from Figure 3.5, the performance of the max-min deterministic robust optimization can be quite sensitive to the choice of α.

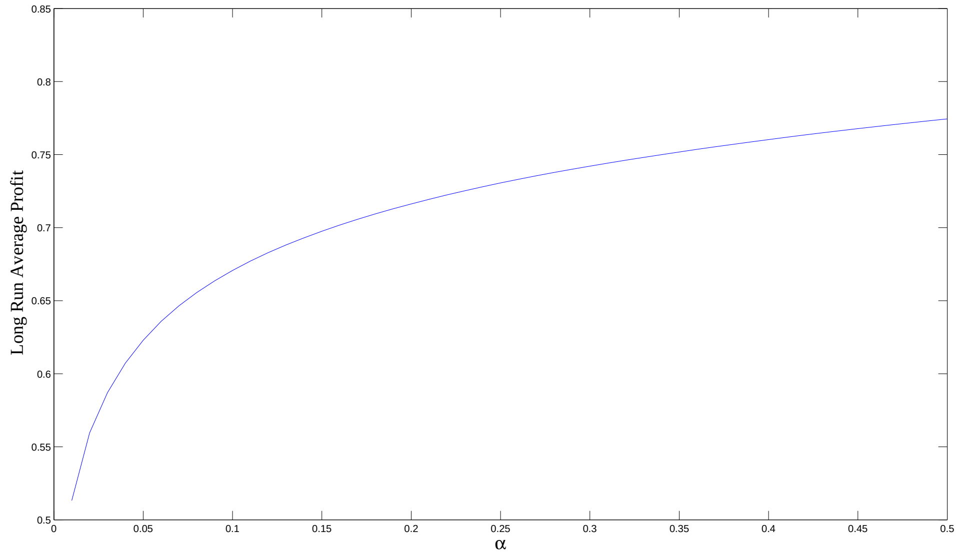

Figure 3.5: Sesitivity of Max-min robust optimization (with deterministic uncertainty set) objective with respect to α .

Max-min with Parametric Uncertainty Set

Suppose the variable parametric uncertainty set in period k is $a_k \leq \theta \leq b_k$. The robust optimization problem in period k is:

$$\max_{y_k \geq 0} \min_{a_k \leq \theta \leq b_k} s\theta \left(1 - \exp\left(-\frac{y_k}{\theta}\right)\right) - cy_k \tag{3.56}$$

As before the inner minimization is monotone in θ, and so it is immediate that

$$y_k^{R:PF} = a_k \log\left(\frac{s}{c}\right), \ k = 1, 2, \ldots, m. \tag{3.57}$$

To implement the solution, we can use confidence interval for θ as defined in (3.45). The robust profit $\psi^{R:PF}$ is:

$$\psi^{R:PF} = \sum_{k=1}^{n} s \min\left\{\frac{2l_k \bar{D}_k^{l_k}}{\chi_{2l_k,\alpha/2}} \log\left(\frac{s}{c}\right), D_k\right\} - c\frac{2l_k \bar{D}_k^{l_k}}{\chi_{2l_k,\alpha/2}}. \tag{3.58}$$

Figure 3.6 shows the performance of max-min robust parametric solution and classical parametric solution for an uncertainty set corresponding to 90% confidence level. The per-

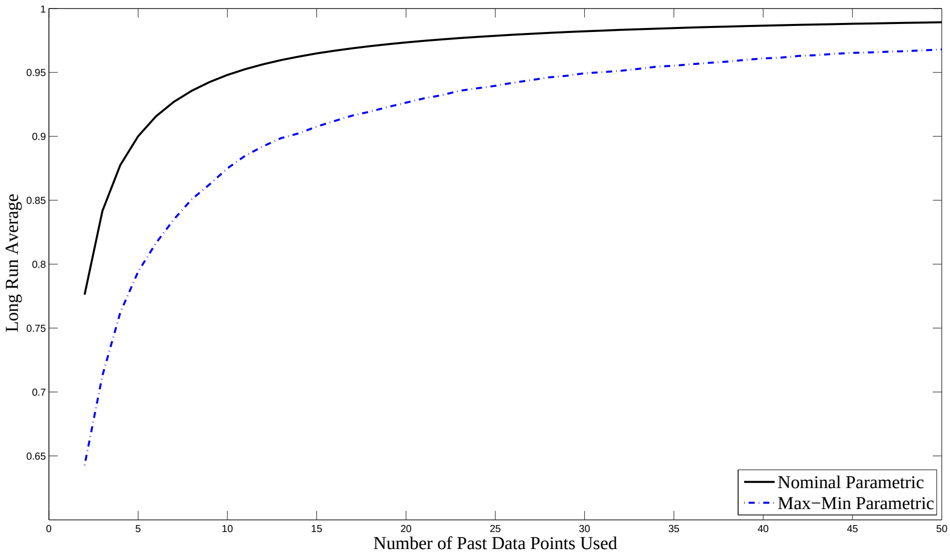

Figure 3.6: Performance of Max-min robust optimization with parametric uncertainty set

formance of robust solution is worse than classical point estimate based parametric solution irrespective of the number of past data points used.

Max-min with Nonparametric Variable Uncertainty Set

The non-parametric uncertainty set for demand D_k in period k is the set of all probability distribution with mean μ_k and variance σ_k^2. It can be shown that maximizing the newsvendor profit is equivalent to minimizing

$$y_k + \frac{s}{c}\mathbb{E}_P\left[D_k - y_k\right]. \tag{3.59}$$

An upper bound on $\mathbb{E}_P\left[D_k - y_k\right]$, as proved in [GM94] is:

$$\mathbb{E}_P\left[D_k - y_k\right] \leq \frac{\left(\sigma_k^2 + (y_k - \mu_k)^2\right)^{1/2} - (y_k - \mu_k)}{2}. \tag{3.60}$$

The order quantity y_k^{NP} that minimizes the upper bound is:

$$y_k^{R:NP} = \mu_k + \frac{\sigma_k}{2}\left(\left(\frac{s-c}{c}\right)^{1/2} - \left(\frac{c}{s-c}\right)^{1/2}\right). \tag{3.61}$$

To actually implement this policy we need estimates of μ_k and σ_k. Let the estimates based on past l_k periods of data be:

$$\hat{\mu}_k^{l_k} = \frac{1}{l_k} \sum_{i=k-l_k}^{k-1} D_k, \quad \text{and} \tag{3.62}$$

$$\hat{\sigma}_k^{l_k} = \frac{1}{l_k - 1} \sum_{i=k-l_k}^{k-1} (D_k - \hat{\mu}_k^{l_k})^{1/2}. \tag{3.63}$$

The implemented policy is:

$$\hat{y}_k^{R:NP} = \hat{\mu}_k^{l_k} + \frac{\hat{\sigma}_k^{l_k}}{2} \left(\left(\frac{s-c}{c} \right)^{1/2} - \left(\frac{c}{s-c} \right)^{1/2} \right). \tag{3.64}$$

The profit $\psi^{R:NP}$ corresponding to policy $\hat{y}_k^{R:NP}$ is:

$$\psi^{R:NP} = \sum_{k=1}^{n} \left(s \min \left\{ \hat{y}_k^{R:NP}, D_k . \right\} - c\hat{y}_k^{R:NP} \right). \tag{3.65}$$

Figure 3.7 compare the performance of max-min robust non-parametric solution and classical non-parametric solution for an uncertainty set corresponding to 90% confidence level. There is not much difference between the performance of robust solution and that of a classical non-parametric solution.

3.4.2 Min-max Regret

Max-min objectives choose the worst model out of all permissible ones. If the model happens to be best one possible then the performance of a robust max-min solution is very bad. There is an opportunity cost if the true model is best one and the solution is too conservative. Min-max regret find a solution by minimizing this opportunity cost. Mathematically, a min-max regret objective is:

$$\min_{\boldsymbol{y} \in \mathcal{Y}} \max_{P \in \mathcal{P}} \mathbb{E}_P \left[\psi(\boldsymbol{y}^*(P), \boldsymbol{X}) - \psi(\boldsymbol{y}, \boldsymbol{X}) \right], \tag{3.66}$$

where $y^*(P)$ is the optimal solution under the probability measure P. The notion of regret is more appealing than worst-case optimization as it sought to achieve a middle ground between optimistic and pessimistic scenarios.

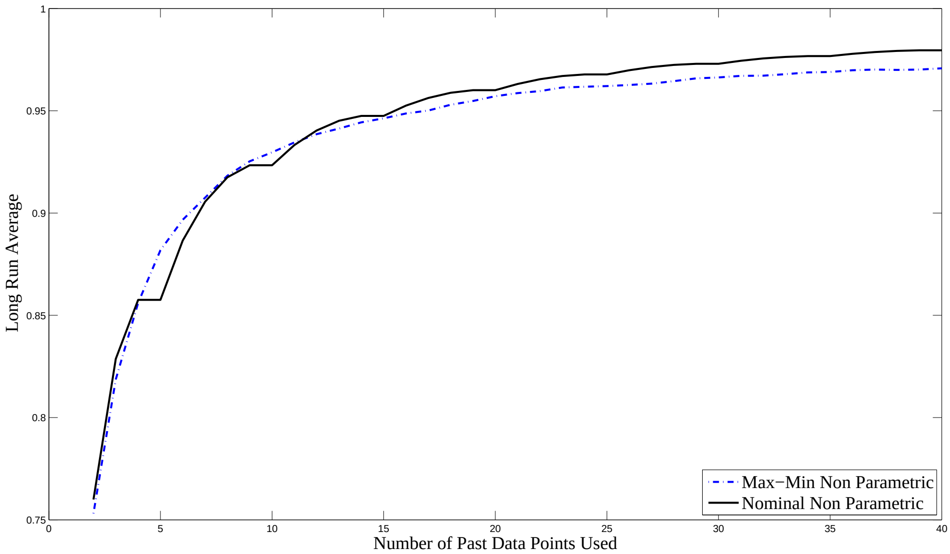

Figure 3.7: Performance of Max-min robust optimization with non-parametric uncertainty set

Min-Max Regret with Deterministic Variable Uncertainty Set

If the demand in period k is known to be deterministically D then the optimal profit is $(s-c)D$. So the regret objective in period k, if the deterministic uncertainty set is $[a_k, b_k]$, is:

$$\min_{y_k} \max_{a_k \le D_k \le b_k} \left((s-c)D_k - s\min\{y_k, D_k\} - cy_k\right). \tag{3.67}$$

For a given y_k the objective is convex in D_k. Hence, the objective in (3.67) can be written as:

$$\min_{y_k} \max\{(s-c)a_k - sa_k - cy_k, (s-c)b_k - sy_k - cy_k\}. \tag{3.68}$$

The solution y_k would be such that:

$$(s-c)a_k - sa_k - cy_k = (s-c)b_k - sy_k - cy_k. \tag{3.69}$$

Hence

$$y_k = a_k + (b_k - a_k)\frac{s-c}{s}. \tag{3.70}$$

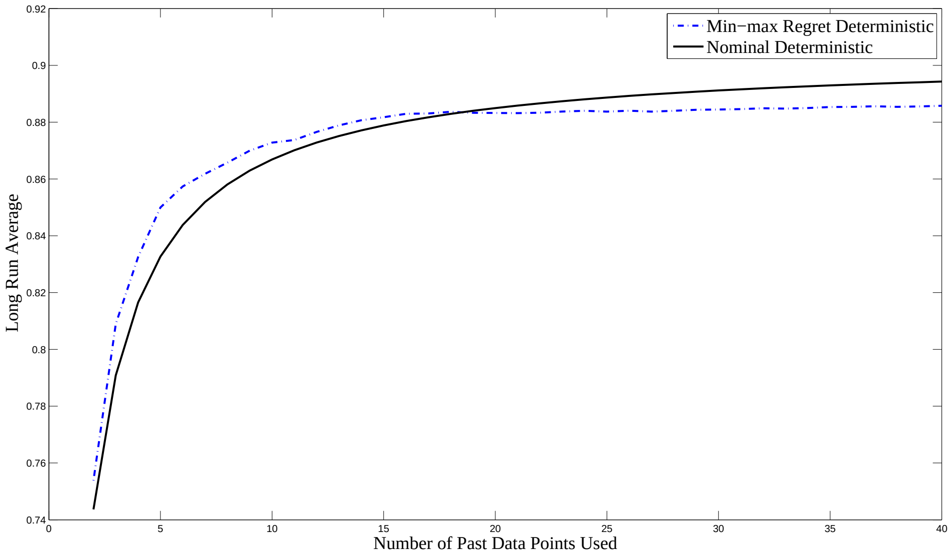

Figure 3.8: Performance of of Min-max regret optimization with deterministic uncertainty set

As we see here the regret solution depends on both a_k and b_k and is also dependent on s and c.

To actually implement this policy if we decide to use values of a_k and b_k as defined in (3.41), the profit, $\psi^{Reg:D}$ is:

$$\psi^{Reg:D} = \sum_{k=1}^{n} \left(s \min \left\{ y_k^{Reg:D}, D_k \right\} - c y_k^{Reg:D} \right), \tag{3.71}$$

where $y_k^{Reg:D}$ is defined by (3.70) with $a_k = \left[\bar{D}_k^{l_k} - z_{\alpha/2} S_k^{l_k} \right]^+$ and $b_k = \bar{D}_k^{l_k} + z_{\alpha/2} S_k^{l_k}$.

Figure 3.8 compares the performance of regret solution to classical solution in case of deterministic uncertainty set. The performance of regret solution is much better than the worst case robust solution and it outperforms the nominal deterministic model for small sample sizes. However, for large sample sizes the classical solution still outperforms the regret solution. Figure 3.9 shows the variation of performance of regret solution with size of deterministic uncertainty set. As compared to the worst case solution the regret solution has a desirable property that it is not too sensitive to the size of the uncertainty set parametrized by α.

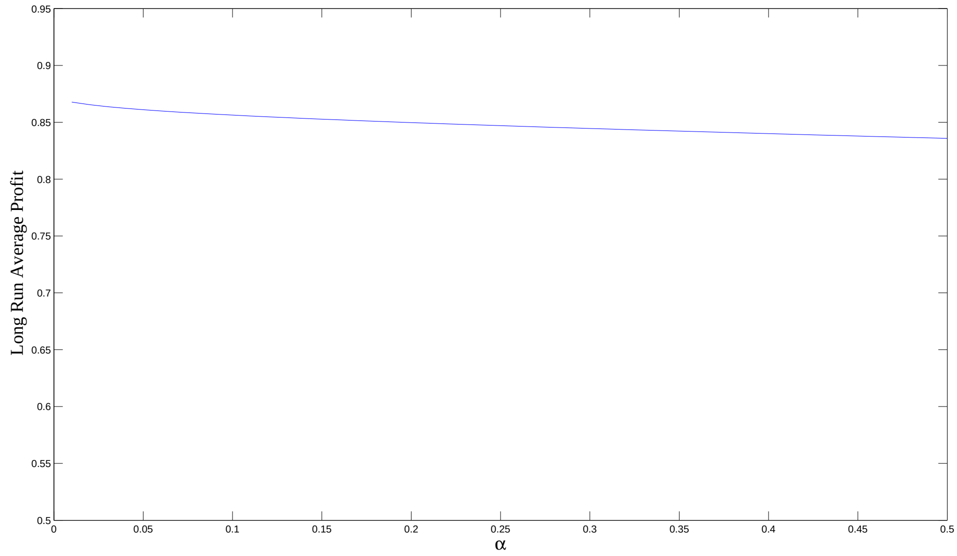

Figure 3.9: Sensitivity of Min-max regret optimization (with deterministic uncertainty set) with α.

Min-Max Regret with Parametric Variable Uncertainty Set

With parameter of exponential lying in $[a_k, b_k]$, the min-max regret problem in period k is

$$\min_{y_k} \max_{a_k \le \theta \le b_k} \mathbb{E}_\theta \left[(s-c)\theta - c\theta \log\left(\frac{s}{c}\right) - s\min\{y_k, D_k\} + cy_k \right] \tag{3.72}$$

which can be written after taking expectation as

$$\min_{y_k} \max_{a_k \le \theta \le b_k} \left[\theta \left(s \exp\left(\frac{-y_k}{\theta}\right) - c\theta(1 + \log(s/c)) \right) + cy_k \right] \tag{3.73}$$

It can be shown that the optimal y_k can be found by equating regrets at $\theta = a_k$ and $\theta = b_k$. The optimum y_k is the solution of the following equation:

$$(b_k - a_k)\frac{c}{s}\left(1 + \log\left(\frac{s}{c}\right)\right) = b_k \exp(-y_k/b_k) - a_k \exp(-y_k/a_k). \tag{3.74}$$

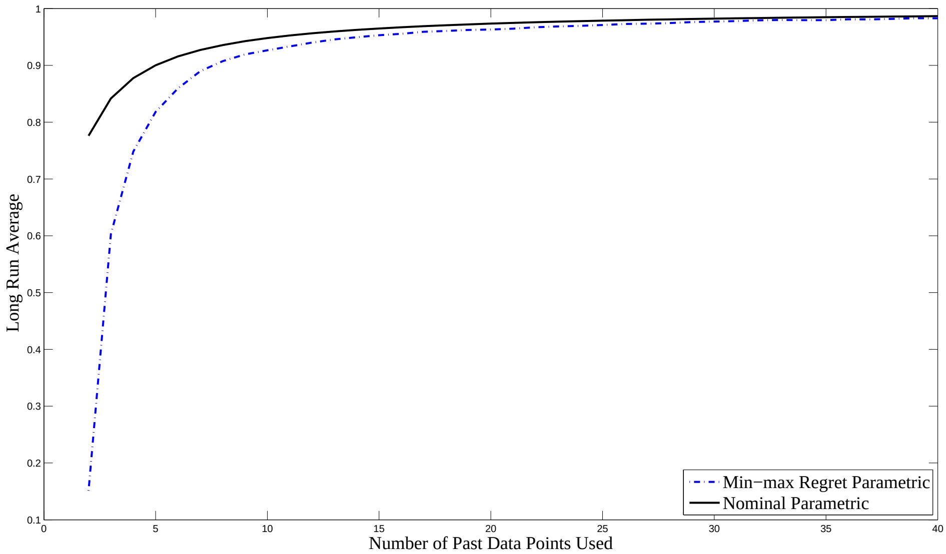

Figure 3.10: Performance of of Min-max regret optimization with parametric uncertainty set

We can use confidence interval for θ as defined in (3.45) for a_k and b_k. The profit function $\psi^{Reg:PF}$ is

$$\psi^{Reg:PF} = \sum_{k=1}^{n} \left(s \min \left\{ y_k^{Reg:PF}, D_k \right\} - c y_k^{Reg:PF} \right), \tag{3.75}$$

with $y_k^{Reg:PF}$ is solution of (3.74) with $a_k = \frac{2l_k \bar{D}_k^{l_k}}{\chi_{2l_k,\alpha/2}}$ and $b_k = \frac{2l_k \bar{D}_k^{l_k}}{\chi_{2l_k,1-\alpha/2}}$.

Figure 3.10 compares the performance of regret solution with classical parametric solution for 90% confidence parametric uncertainty set. Even though the performance of regret solution is better than the worst case solution, it is still worse than the classical parametric solution for all sample sizes.

Max-min Regret with Nonparametric Variable Uncertainty Set

We consider the the non-parametric uncertainty set for demand D_k in period k is the set of all probability distribution with mean μ_k. It can be shown (See [PR08]) that order quantity that minimize the maximum regret is:

$$y_k^{Reg:NP} = \begin{cases} \mu_k \left(1 - \frac{c}{s}\right) & \text{if } \frac{1}{2} \leq \frac{c}{s} \\ \frac{\mu_k}{4\frac{c}{s}} & \text{if } \frac{1}{2} > \frac{c}{s} \end{cases} \tag{3.76}$$

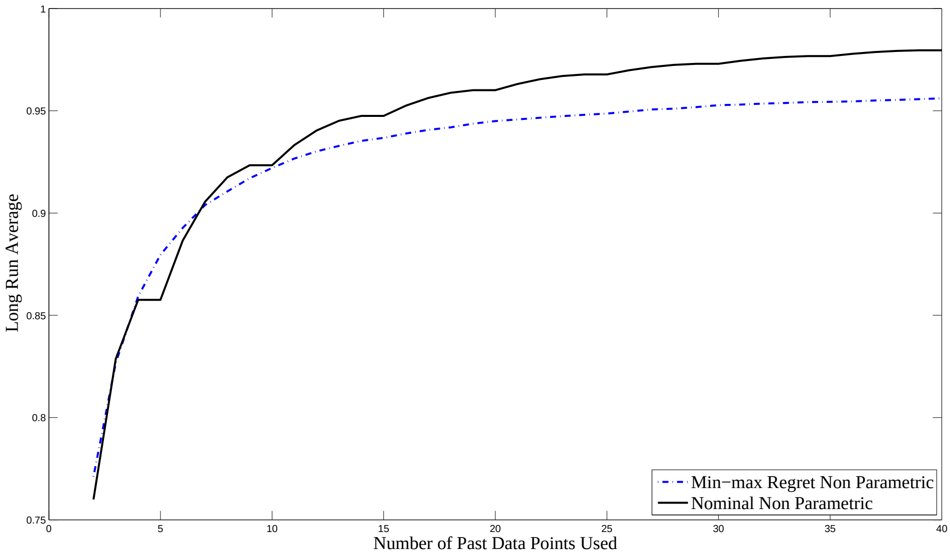

Figure 3.11: Performance of Min-max Regret optimization with non-parametric uncertainty set

The implemented $\tilde{y}_k^{Reg:NP}$ use an estimate of μ_k as in (3.62). The profit $\psi^{Reg:NP}$ corresponding to policy $\hat{y}_k^{Reg:NP}$ is:

$$\psi^{R:NP} = \sum_{k=1}^{n} \left(s \min \left\{ \hat{y}_k^{Reg:NP}, D_k \right\} - c\hat{y}_k^{Reg:NP} \right). \tag{3.77}$$

From Figure 3.11, the performance of non-parametric solution is worse than the classical non-parametric solution based on empirical distribution.

3.4.3 Max-min Competitive Ratio

Max-min Competitive ratio is a benchmark objective similar in spirit to max-min regret. Instead of looking at the difference in the objective values, we look at the ratios. The optimization problem is:

$$\max_{\boldsymbol{y}} \min_{P \in \mathcal{P}} \frac{\mathbb{E}_P \left[\psi(\boldsymbol{y}, \boldsymbol{X} \right]}{\mathbb{E}_P \left[\psi(\boldsymbol{y}^*(P), \boldsymbol{X}) \right]} \tag{3.78}$$

A possible advantage of competitive ratio over regret is the objective in (3.78) is relatively agnostic to the scale of the problem.

Max-min Competitive Ratio with Deterministic Variable Uncertainty Set

The competitive ratio optimization problem in period k (assuming $a_k > 0$) is

$$\max_{y_k} \min_{a_k \le D_k \le b_k} \frac{s \min\{y_k, D_k\} - cy_k}{(s-c)D_k}. \tag{3.79}$$

The optimal y_k can be found by solving the following:

$$\frac{sa_k - cy_k}{a_k} = \frac{sy_k - cy_k}{b_k}, \tag{3.80}$$

from which we obtain

$$y_k = \frac{sa_k b_k}{c(b_k - a_k) + sa_k}. \tag{3.81}$$

If we use values of a_k and b_k as defined in (3.41), the profit, $\psi^{C:D}$ is:

$$\psi^{C:D} = \sum_{k=1}^{n} \left(s \min\left\{ y_k^{C:D}, D_k \right\} - c y_k^{C:D} \right), \tag{3.82}$$

where $y_k^{C:D}$ is defined by (3.81) with $a_k = \max\left\{ \bar{D}_k^{l_k} - z_{\alpha/2} S_k^{l_k}, \epsilon \right\}$ and $b_k = \bar{D}_k^{l_k} + z_{\alpha/2} S_k^{l_k}$. $\epsilon > 0$ is small parameter added to make $a_k > 0$.

Figure 3.12 compares the performance of the solution derived from competitive ratio objective (uncertainty set constructed using 90% confidence level) to the performance of the classical deterministic solution. The competitive ratio solution outperforms the classical deterministic solution for nearly all sample sizes. Hence, as a robust objective it works much better than regret and worst case optimization objectives for the newsvendor problem. In addition, as we see from Figure 3.13 the performance of the competitive ratio solution is not much sensitive to the size of uncertainty set.

Max-min Competitive Ratio with Parametric Variable Uncertainty Set

As before the max-min competitive ratio problem in period k with variable parametric uncertainty is

$$\max_{y_k} \min_{a_k \le \theta \le b_k} \frac{s\theta\left(1 - \exp\left(-\frac{y_k}{\theta}\right)\right) - cy_k}{\theta\left((s-c) - \log\left(\frac{s}{c}\right)\right)}. \tag{3.83}$$

For a given y_k the above objective has maximum at $\theta = \frac{y_k}{\ln(s/c)}$, decreasing on either side of it. In addition, the function is concave in y_k. It can be proved that the optimal policy y_k is

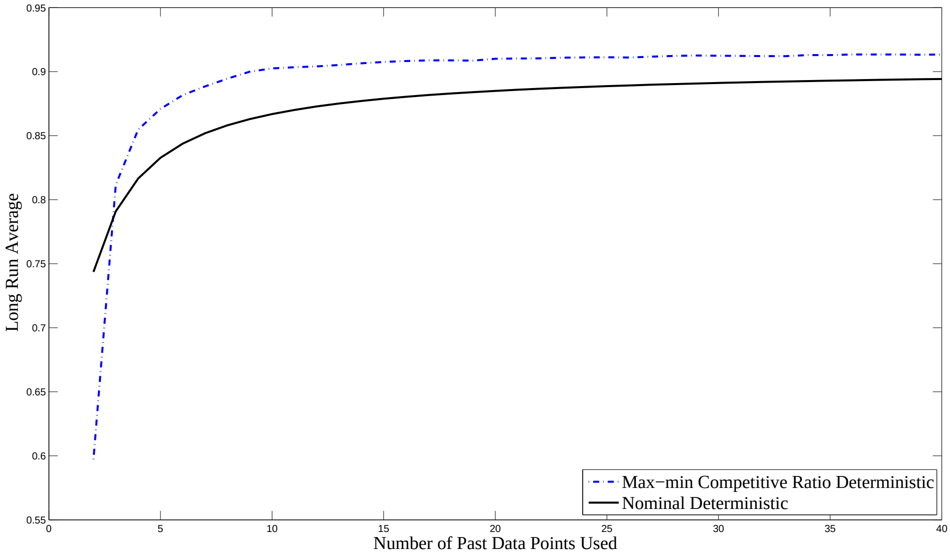

Figure 3.12: Performance of Max-min competitive ratio optimization with deterministic uncertainty set

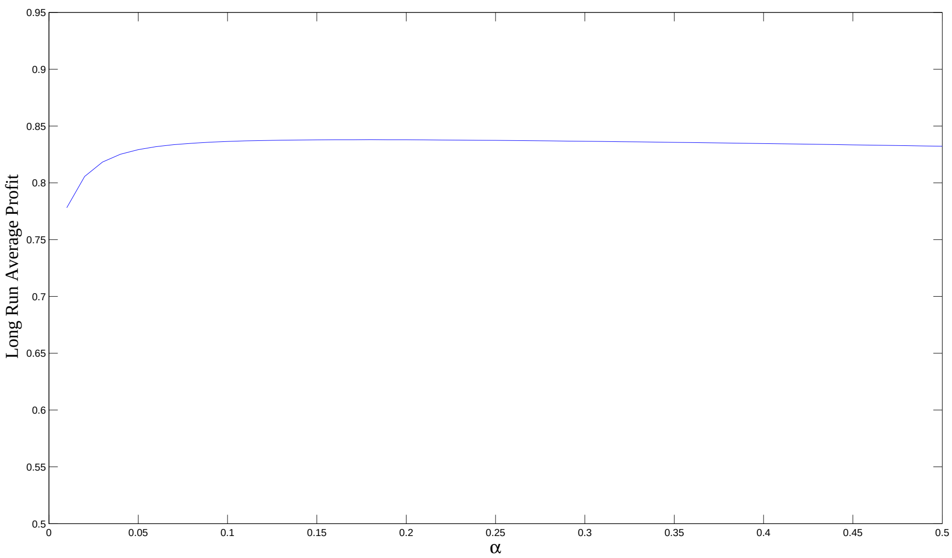

Figure 3.13: Sensitivity of Max-min competitive ratio optimization (with deterministic uncertainty set) with α.

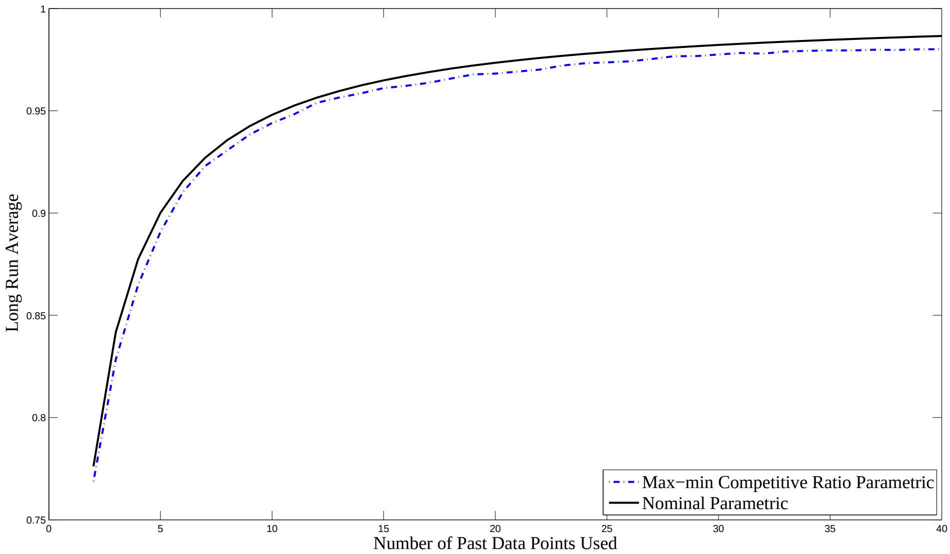

Figure 3.14: Performance of Max-min competitive ratio optimization with parametric uncertainty set

given by solution of the following:

$$\frac{sa_k\left(1-\exp\left(-\frac{y_k}{a_k}\right)\right)-cy_k}{a_k}=\frac{sb_k\left(1-\exp\left(-\frac{y_k}{b_k}\right)\right)-cy_k}{b_k}. \tag{3.84}$$

We can implement the policy by solving for y_k with a_k and b_k given by (3.45). If the solution is $y_k^{C:PF}$, the profit function $\psi^{C:PF}$ is:

$$\psi^{C:PF}=\sum_{k=1}^{n}\left(s\min\left\{y_k^{C:PF},D_k\right\}-cy_k^{C:PF}\right). \tag{3.85}$$

Figure 3.14 compares the performance of the competitive ratio solution with parametric uncertainty set to the solution of the classical parametric solution. The performance of competitive ratio solution is close to the classical parametric solution but is still uniformly worse than the classical solution for all sample sizes. This is in contrast to the competitive ratio with deterministic uncertainty set, where the robust solution outperforms the classical deterministic solution.

The comparison of various robust optimization models when uncertainty set is calculated from past data shows that most of the time performance of robust models is even worse than their non-robust counterparts. Next we discuss some objective learning approaches that integrate optimization and learning.

3.5 Learning

3.5.1 Reinforcement Learning

Reinforcement learning as a learning paradigm falls between supervised and unsupervised learning. Learning is done through trial and error by providing reinforcement or reward in response to an action. The method is particularly suitable for learning in dynamic environment without making too many assumptions about the underlying process or environment. The basic premise of reinforcement learning is to learn the best action in a given "state" by *exploiting* the current best action and also by randomly *exploring* other actions. The "state" can be understood to be a finite dimensional vector which is an input to the system. An example of a state is the initial inventory at the start of a period in inventory control problem. The system tries to learn the "value" of the state, which is the reward obtained by implementing an optimal policy through a reinforcement learning algorithm. Some of the popular reinforcement learning algorithms are TD(λ) [Sut88], Q learning [Wat89], and Dyna [Sut90]. For a survey of reinforcement learning algorithms see [SB98] and [KLM96].

The newsvendor problem, which is an example of sequential decision making, can be formulated as a multiple arm bandit problem, which is a special case of reinforcement learning with single state but multiple actions. Although the action (order quantity) space is continuously valued, we can get a good approximation by discretization the action space into finitely many values then applying an algorithm such as UCB1 [ACBF02] for multi-armed bandit on the discretized problem. A better approach would be to apply continuum arm bandit algorithm [Kle04], as the newsvendor profit function is Lipschitz continuous in order quantity. An advantage of multi-arm bandit approach is that we can theoretically bound the cumulative regret of the algorithm. On the other hand the small sample performance of multi-arm bandit algorithm does not utilize much structural and distributional information about the problem in hand.

3.5.2 Objective Bayesian

The main criticism of subjective Bayesian methods is subjectivity in choosing the prior for the parameters of distribution of stochastic process. It is often difficult to find an "expert" who can provide a good subjective prior. As we have seen from Figure 3.1 the solution does depend heavily on the choice of prior. A possible alternative is to use non-informative/objective priors. A non-informative prior can informally be defined as the one which favors no particular value of parameter over others. One particular way to define non-informative is to use Jefferys prior [Jef46], which is to use

$$f_\Theta(\theta) \propto (I(\theta))^{1/2} \tag{3.86}$$

as non-informative prior, where $I(\theta)$ is the expected Fisher information. Under some commonly satisfied assumptions [Ber85] $I(\theta)$ is given by

$$I(\theta) = -\mathbb{E}_\theta \frac{\partial^2 \log P_\theta}{\partial \theta^2}. \tag{3.87}$$

For an exponential distribution with mean θ, the expected Fisher information $I(\theta)$ is $\left(\frac{1}{\theta}\right)^2$, and a non-informative prior is:

$$f_\Theta(\theta) = \frac{1}{\theta}, \qquad \theta > 0.. \tag{3.88}$$

Note that the prior (3.88) is not a proper probability density over $(0, \infty)$ as $\int_{(0,\infty)} \frac{1}{\theta} d\theta = \infty$. However, we may still use $f_\Theta(\theta)$ as a probability density in Bayes' rule if the posterior is a proper probability density.

For the newsvendor problem, if in period k we use past l_k periods of data to make the decision then the posterior distribution using prior (3.88) is

$$f_\Theta^k(\theta) = \frac{l_k \bar{D}_k^{l_k}}{(l_k - 1)! \theta^{l_k+1}} \exp\left(-\frac{l_k \bar{D}_k^{l_k}}{\theta}\right). \tag{3.89}$$

Hence the marginal distribution of D_k is

$$f_{D_k}(x) = \int_\Theta \frac{\left(l_k \bar{D}_k^{l_k}\right)^{l_k}}{(l_k - 1)! \theta^{l_k+2}} \exp\left(-\frac{l_k \bar{D}_k^{l_k} + x}{\theta}\right) d\theta. \tag{3.90}$$

which is equal to

$$f_{D_k}(x) = \frac{l_k \left(l_k \bar{D}_k^{l_k}\right)^{l_k}}{\left(l_k \bar{D}_k^{l_k} + x\right)^{l_k+1}}. \tag{3.91}$$

Hence

$$\bar{F}_{D_k}(x) = \left(\frac{l_k \bar{D}_k^{l_k}}{l_k \bar{D}_k^{l_k} + x}\right)^{l_k}. \tag{3.92}$$

Therefore, the optimal order quantity $y^{OB} = \bar{F}_{D_k}^{-1}(c/s)$ is:

$$y^{OB} = D_k^{l_k} l_k \left((s/c)^{1/l_k} - 1\right). \tag{3.93}$$

Objective Bayesian learning provides an attractive way to make decisions in the presence of model uncertainty. Later in Chapter 4 we will see this is (in a way) the best we can do.

However, objective Bayesian can also be criticized similarly to subjective Bayesian as there is no unique way of defining a non-informative prior. For example, for Normal distribution with mean μ and variance σ^2, Jefferys prior is $\frac{1}{\sigma^2}$. Another commonly used non-informative prior is $\frac{1}{\sigma}$. It is not clear which prior to be used for a particular problem. In addition, similar to subjective Bayesian method the prior should be chosen independent of objective function.

3.5.3 Classical Decision Theory

In classical (non-Bayesian) decision theory a non-randomized decision rule $\boldsymbol{y}$ is a mapping from the available information to the policy space. So the decision at time k, y_k is directly a function of the information set $\mathcal{I}_k$. We denote the dependency of $\boldsymbol{y}$ on information set by $\boldsymbol{y}(\mathcal{I})$. A classical decision theorists seeks to evaluate the "loss" for every value of the parameter of the underlying stochastic process. The loss is measured by using a *risk function*. In the setup where profit is maximized we can use the negative of the profit function as risk. Equivalently we can focus on

$$\max_{\boldsymbol{y}} \mathbb{E}_\theta \left[\psi(\boldsymbol{y}(\mathcal{I}), \boldsymbol{X}(\theta)\right]. \tag{3.94}$$

As the frequentist risk or profit function is a function of θ, it is not clear what maximizing (3.94) actually means. A partial ordering of decision rules can be formed by utilizing the following notion:

Definition 7. *A decision rule $\boldsymbol{y}(\mathcal{I})$ is better than $\boldsymbol{y}'(\mathcal{I})$ if*

$$\mathbb{E}_\theta \left[\psi(\boldsymbol{y}(\mathcal{I}), \boldsymbol{X}(\theta)\right] \geq \mathbb{E}_\theta \left[\psi(\boldsymbol{y}'(\mathcal{I}), \boldsymbol{X}(\theta)\right] \qquad \forall\, \theta \tag{3.95}$$

with strict inequality for at least one θ.

Definition 8. *A decision rule is* admissible *if there exists no better decision rule as defined in Definition 7.*

Naturally it makes sense to search only for decision rules which are admissible. However, the class of admissible decision rules is very large. If we fix a particular value of θ, say θ_0 and let $\tilde{\boldsymbol{y}}$ be the optimal decision for θ_0 regardless of past data values or information, then it is likely that $\tilde{\boldsymbol{y}}$ is admissible. However, such a decision rule is likely to perform quite bad for other values of θ. On the other hand we can define the concept of an optimal decision rule as follows:

Definition 9. *A policy* $\boldsymbol{y}$ *is optimal, if*

$$\mathbb{E}_\theta\left[\psi(\boldsymbol{y}(\mathcal{I}), \boldsymbol{X}(\theta))\right] \geq \mathbb{E}_\theta[[\psi(\boldsymbol{y'}(\mathcal{I}), \boldsymbol{X}(\theta))]\,, \forall\, \theta, \tag{3.96}$$

for all possible $\boldsymbol{y'}$*.*

Such a concept, though attractive, is quite restrictive. Indeed the class of problems where one can find a decision rule in the sense of (3.96) is quite small. Therefore, there is a need to define alternative concepts of choosing decision rules. In classical decision theory, there are three major methods [Ber85] of choosing decision rules. First is *Bayes Risk Principle* where the risk or profit function is integrated over θ using a prior on θ. The method is not too different from subjective Bayesian principle discussed in Section 3.1.2, and can be criticized similarly for subjectivity in choosing the prior distribution. Second is *Invariance Principle* which informally states that a decision rule should not depend on the unit of measurement used or any other such arbitrary structure. There is a direct correspondence of such rules with method of using non-informative priors and hence we will not discuss it here. Third is the *Max-min Principle*. The aim is to choose a decision rule that maximize the profit assuming the worst possible parameter of stochastic process. Such a concept is similar to robust optimization, the difference being the policy is a function of past data/information.

3.6 Discussion

In this chapter, we have studied the effect of model uncertainty on various classical modeling approaches. We studied many popular robust optimization approaches and compared the performance to their non-robust counterparts. If the performance is compared on the basis of next period expected profit, then most of the time performance of robust solutions is even worse than the classical non-robust solutions. In Chapter 4, we introduce generalized operational statistics approach, which is a parametric approach that guarantees a better solution than a classical solution over a set of parameter.

In Section 3.2.2, we also discussed the effect of using wrong model class and wrong statistical assumption. When confidence in a particular model class is low it is desirable to use a non-parametric method such as reinforcement learning. In Chapter 6 we introduce a new non-parametric learning method, which utilizes the structural information embedded in a problem to improve small sample performance.

Chapter 4

Operational Statistics

In Chapter 3, we discussed different approaches for modeling uncertainty with learning. In almost all these approaches the learning and optimizations tasks are separated. First the parameters of the model are statistically estimated, and then optimization is done on the resulting model assuming that the estimated parameters are true ones. Operational statistics (OS) is an integrated learning and optimization approach to model uncertainty. The OS approach was first introduced in [LS05] and further explored in [CSS08]. The papers [LS05] and [CSS08] have successfully applied operational statistics to the parametric setting where the unknown parameters are either scale or shift parameters. In this chapter we generalize the operational statistics approach and in Chapter 5, we apply the approach to mean variance portfolio optimization problem. Below we present a brief introduction to the operational statistics approach as applied to the newsvendor problem [LS05]. We then present a generalization of this apporach.

4.1 Operational Statistics

Consider the newsvendor problem as defined in Chapter 3. The expected profit when the order quantities and the realized demands in first n periods are $y_1, \ldots, y_n$ and $D_1, \ldots, D_n$ respectively is:

$$\phi(\mathbf{y}) = \sum_{k=1}^{n} \mathbb{E}\left[\{s \min\{y_k, D_k\} - cy_k\}\right]. \tag{4.1}$$

If we know the distribution of the demand then it is well known that the optimal order quantity in every period is $\bar{F}^1\left(\frac{c}{s}\right)$ where F is the distribution function of the demand. Now consider the case when the demand is exponentially distributed with unknown parameter θ. If parameter θ is known then the optimal order quantity is $\theta \ln\left(\frac{c}{s}\right)$. When θ is unknown one particular method that is frequently used in the literature is to replace θ by its point estimate at every time instant. When θ is a scale parameter a point estimate of θ for period

k using past l_k periods of demand data is:

$$\tilde{\theta}_k = \bar{D}_k^{l_k} = \frac{1}{l_k} \sum_{i=k-l_k}^{k-1} D_i. \tag{4.2}$$

Hence the order quantity $\hat{y}_k$, using the estimate $\tilde{\theta}_k$ is:

$$\hat{y}_k = \bar{D}_k^{l_k} \ln\left(\frac{s}{c}\right). \tag{4.3}$$

If we use the order quantity $\hat{y}_k$, then using equation (4.1), the a priori expected profit ϕ_{sm} as a function of θ is:

$$\begin{aligned} \phi_{sm}(\theta) = \mathbb{E}\psi_{sm}(\boldsymbol{D}) &= \sum_{k=1}^{n} \mathbb{E}\left[\left\{s \min\left\{\bar{D}_k^{l_k} \ln(\left(\frac{s}{c}\right), D_k\right\} - c\bar{D}_k^{l_k} \ln\left(\frac{s}{c}\right)\right\}\right] \\ &= \sum_{k=1}^{n} c\theta \left[\frac{s}{c} - \frac{s}{c}\left(\frac{l_k}{l_k + \ln(s/c)}\right)^{l_k} - \ln\left(\frac{c}{s}\right)\right]. \end{aligned} \tag{4.4}$$

In operational statistics an order quantity, rather than the parameters of the distribution, is directly estimated from the data. In this sense, the approach is similar to the policy gradient and direct reinforcement methods mentioned in the reinforcement learning literature (see Baxter and Bartlett [BB01] and Moody and Saffell [MS01]. However, in operational statistics the order quantity is estimated in such a way that the a priori expected profit is maximized, rather than doing trial and error and fine-tuning the parameters using gradients in policy space. The latter approach requires extensive amount of data before it performs well and is unsuitable for most of the problems where past data is not readily available. Even if the data is available, it is not necessary that the statistical assumptions (such as i.i.d. demands) are valid over a long run of time.

The idea of operational statistics is as follows: we consider order quantities to be statistics of the data $\{D_1, D_2, \ldots, D_n\}$ within some acceptable class parametrized by some optimization variables, say $\mathbf{z}$. This statistics is called operational statistics as the statistics is driven by the operational problem, not by the estimation problem. One then finds an optimal operational statistics within the class by maximizing the expected a priori profit over the variables $\mathbf{z}$. Specifically for the problem in hand, where the demand is exponential with unknown scale parameter θ, the order quantity is chosen to be of the form

$$\tilde{y}_k(z_k) = z_k \bar{D}_k^{l_k}, \qquad z \geq 0. \tag{4.5}$$

The a priori expected profit $\phi_{os}(\theta)$ using the estimate $\tilde{y}_k$ is

$$\phi_{os}(\theta) = \sum_{k=1}^{n} \mathbb{E}\left[\left\{s \min\left\{z_k \bar{D}_k^{l_k}, D_k\right\} - c z_k \bar{D}_k^{l_k}\right\}\right] = \sum_{k=1}^{n} c\theta \left[\frac{s}{c} - \frac{s}{c}\left(\frac{l_k}{l_k + z_k)}\right)^{l_k} - z_k\right]. \tag{4.6}$$

Optimizing over z_k we get

$$z_k^* = l_k \left(\left(\frac{s}{c}\right)^{1/(l_k+1)} - 1\right). \tag{4.7}$$

The operational statistics order quantity is

$$\tilde{y_k} = l_k \left(\left(\frac{s}{c}\right)^{1/(l_k+1)} - 1\right) \bar{D}_k^{l_k}. \tag{4.8}$$

Realized profit ψ_{OS} using operational statistics is

$$\psi_{OS}(\boldsymbol{D}) = \sum_{k=1}^{n} \left[\left\{s \min\left\{l_k \left(\left(\frac{s}{c}\right)^{1/(l_k+1)} - 1\right) \bar{D}_k^{l_k}, D_k\right\} - c l_k \left(\left(\frac{s}{c}\right)^{1/(l_k+1)} - 1\right) \bar{D}_k^{l_k}\right\}\right]. \tag{4.9}$$

Consider the situation where the underlying process $\{D_k\}$ is locally stationary. At every time instant only last m periods of data can be considered approximately i.i.d,. and therefore we use $l_k = m$ for every k. Long run average profit of the operational statistics strategy would be

$$\begin{aligned}\lim_{n\to\infty} \frac{\psi_{OS}(\boldsymbol{D})}{n} &= \mathbb{E}\left[\left\{s \min\left\{m \left(\left(\frac{s}{c}\right)^{1/(m+1)} - 1\right) \bar{D}_k^m, D_k\right\} - cm \left(\left(\frac{s}{c}\right)^{1/(m+1)} - 1\right) \bar{D}_k^m\right\}\right] \\ &= \theta \left[\left(s - s\left(\frac{c}{s}\right)^{m/(m+1)}\right) - cm\left(\left(\frac{s}{c}\right)^{1/(m+1)} - 1\right)\right].\end{aligned} \tag{4.10}$$

Compare (4.10) to the long run average cost of using a trailing m period sample mean strategy

$$\lim_{n\to\infty} \frac{\psi_{OS}(\boldsymbol{D})}{n} = \theta\left[s - s\left(\frac{m}{m + \ln(s/c)}\right)^m - c\ln\left(\frac{c}{s}\right)\right]. \tag{4.11}$$

Since the operational statistics order quantity (4.8) maximizes the a priori expected profit over all order quantities of the form $z\bar{D}_k^{l_k}, z \geq 0$, the expected (or long run average) profit obtained from the operational statistics approach (4.10) is higher than the profit obtained from the sample mean based approach (4.11) for all choices of m. On the other hand, as seen in Chapter 3, none of the robust optimization approaches guarantees a uniformly better(in a priori expected profit sense) solution than a simple sample mean based approach. As we see from a sample simulation shown in Figure (4.1) for parameters $s = 2, c = 1$ and $\theta = 1$, the improvement can be quite significant for small values of m.

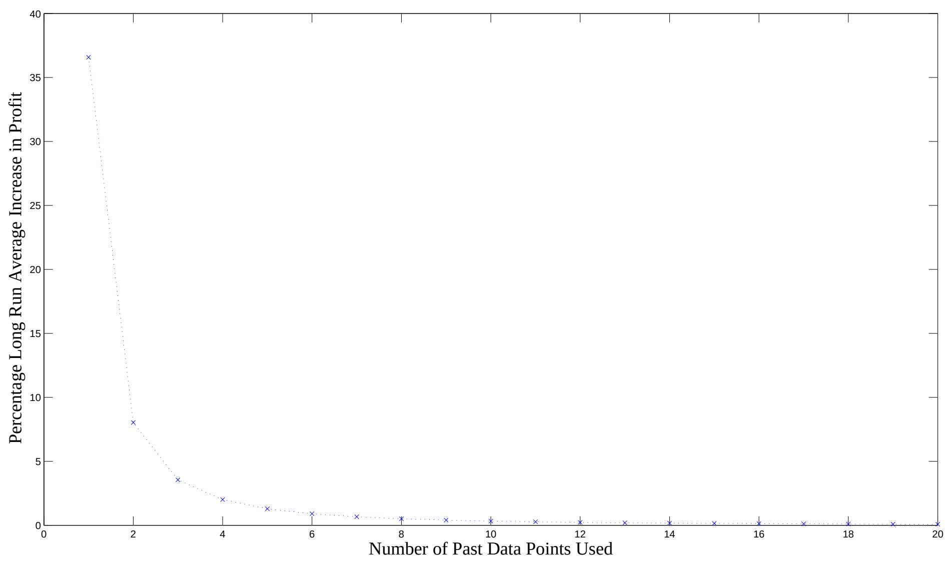

Figure 4.1: Operational statistics policy is uniformly better than the sample mean based policy for newsvendor problem

Note that in equation (4.7), the optimal z_k is independent of θ. This is because the unknown parameter θ acts as a scaling parameter in the profit function (4.6). It is shown in [LS05], when the unknown parameter is a scale parameter (e.g. mean in exponential distribution or normal distribution with unknown mean) and the class of operational statistics is linear in sample mean, the optimal operational statistics is independent of the scaling parameter for newsvendor problem. The paper [CSS08] extended the result of [LS05] to a class of homogeneous order one function and two parameter families (scale and shift). They also showed a connection of operational statistics to objective Bayesian decision theory. However, the class of problems where an optimal operational statistics is independent of the underlying unknown parameter is very limited or difficult to find.

4.1.1 Generalized Operational Statistics

A defining feature of operational statistics, as mentioned above, is that it guarantees a better solution than the classical estimation based policy. However, it is either very hard or impossible to find a statistics such that the corresponding profit is uniformly better than the profit of a classical estimation based policy over the full range of parameter. Suppose in addition the decision maker knows for sure that the unknown parameter lies in a known uncertainty set Θ. Our objective is to improve the performance of classical policy over the range Θ. In other words, we want to ensure that the performance of operational statistics policy is better than the classical estimation based policy over the set Θ. A generalized version of operational statistics is implemented as follows:

Suppose we are interested in maximizing the function $\mathbb{E}_\theta \psi(y, \boldsymbol{D})$ over y. If the value of θ is known an optimal policy is $y(\theta)$. In the absence of true knowledge about θ, a classical estimation based policy is $y(\hat{\theta})$, where $\hat{\theta}$ is an estimate of θ. In addition, the decision maker knows or strongly believes that the parameter θ lies in a set Θ. We choose a set of functions $\mathcal{F}$, such that $\mathcal{F}$ defines a class of operational statistics. Each function $f \in \mathcal{F}$ defines a mapping from past data $\boldsymbol{D}$ to a decision y. The function class $\mathcal{F}$ is such that the classical estimation based decision rule $y(\hat{\theta})$ belongs to $\mathcal{F}$. In addition we choose a representative parameter $\tilde{\theta}$ and solve the following optimization problem:

Operational Statistics Formulation 1

$$\begin{aligned} & \max_{f \in \mathcal{F}} \mathbb{E}_{\tilde{\theta}} \psi(f(\boldsymbol{D}), \boldsymbol{D}) \\ & \text{such that} \\ & \mathbb{E}_\theta \psi(f(\boldsymbol{D}), \boldsymbol{D}) \geq \mathbb{E}_\theta \psi(y(\hat{\theta}), \boldsymbol{D}) \qquad \forall \theta \in \Theta \end{aligned} \tag{4.12}$$

The representative parameter $\tilde{theta}$ is subjective knowledge of the decision maker independent from the data. Thus, the decision maker wants to utilize her subjective information, but because she is not sure about her subjective knowledge, wants a solution that is guaranteed to perform better than the classical solution over the acceptable range of parameters. As $y(\hat{\theta})$ belongs to $\mathcal{F}$, $y(\hat{\theta})$ is a feasible solution of (4.12) and hence the optimization problem (4.12) is feasible. Also even though the choice of $\tilde{\theta}$ is subjective, the constraints in the optimization problem ensure that the operational statistics solution is better than the classical solution over the set Θ, and hence ensures robustness of the solution with respect to the parameter $\tilde{\theta}$. Instead of using a single parameter $\tilde{\theta}$, we can also use a subjective prior on θ. Let P_θ be the choice of the prior. Following is the version of operational statistics problem with subjective prior:

Operational Statistics Formulation 2

$$\begin{aligned} & \max_{f \in \mathcal{F}} \int_\Theta \mathbb{E}_{\tilde{\theta}} \psi(f(\boldsymbol{D}), \boldsymbol{D}) P_\theta \\ & \text{such that} \\ & \mathbb{E}_\theta \psi(f(\boldsymbol{D}), \boldsymbol{D}) \geq \mathbb{E}_\theta \psi(y(\hat{\theta}), \boldsymbol{D}) \qquad \forall \theta \in \Theta \end{aligned} \tag{4.13}$$

As the operational statistics solution is guaranteed to perform better than classical solution, and a classical solution converges to the true solution when $\hat{\theta} \to \infty$, the operational statistics solution would also converge to the true solution. The uncertainty set Θ, on the other hand, is held fixed and is not updated in response to the past data. It is also possible to use some other benchmark instead of classical estimation policy in the operational statistics formulation.

Chapter 5

Application of Operational Statistics to Portfolio Optimization

In this chapter, we consider an important application of operational statistics approach defined in Chapter 4 in mean variance portfolio optimization problem.

5.1 Mean-Variance Portfolio Optimization

We consider the problem of finding a mean variance optimal portfolio. The investor's problem is to find an optimal allocation to m-risky assets and a risk-free asset. Let $\boldsymbol{\mu}$ be the vector of mean excess returns (return - risk free rate) and Σ be the covariance matrix. We assume that the portfolio returns are normally distributed. The investor wishes to maximize the following mean variance utility function:

$$\max_{\boldsymbol{\pi}} \boldsymbol{\pi}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{\pi}\Sigma\boldsymbol{\pi}, \tag{5.1}$$

where $\boldsymbol{\pi}$ is the vector of portfolio weights and γ is the risk aversion factor. The fraction $1-\boldsymbol{\pi}'\boldsymbol{e}$ is invested in the risk free asset. Note that we do not have any short selling constraints.

The problem can be solved analytically if we have exact knowledge of parameters $\boldsymbol{\mu}$ and Σ. In fact, the solution is $\boldsymbol{\pi} = \frac{1}{\gamma}\Sigma^{-1}\boldsymbol{\mu}$. In reality, obtaining an accurate estimate of $\boldsymbol{\mu}$ is quite difficult. For example, if the return of a stock is generated from a normal distribution with mean 20% and volatility 20%, it would take around 384 years of data to estimate the return within an error of 10% with 95% confidence as can be seen from Figure 5.1 To see how uncertainty in the parameters affects the utility function in (5.1) consider the following example. There is one risky asset and one risk free asset in the portfolio. Return is normally distributed with mean μ and known variance σ^2. The investor solves the maximum utility problem corresponding to (5.1) and estimates the fraction of wealth invested in the risky asset to be $\pi = \frac{1}{\gamma\sigma^2}\mu$. Since the parameter μ is unknown, she estimates it using past n periods of return data. The estimate $\hat{\mu}$, which is a maximum likelihood estimate, is the average of last

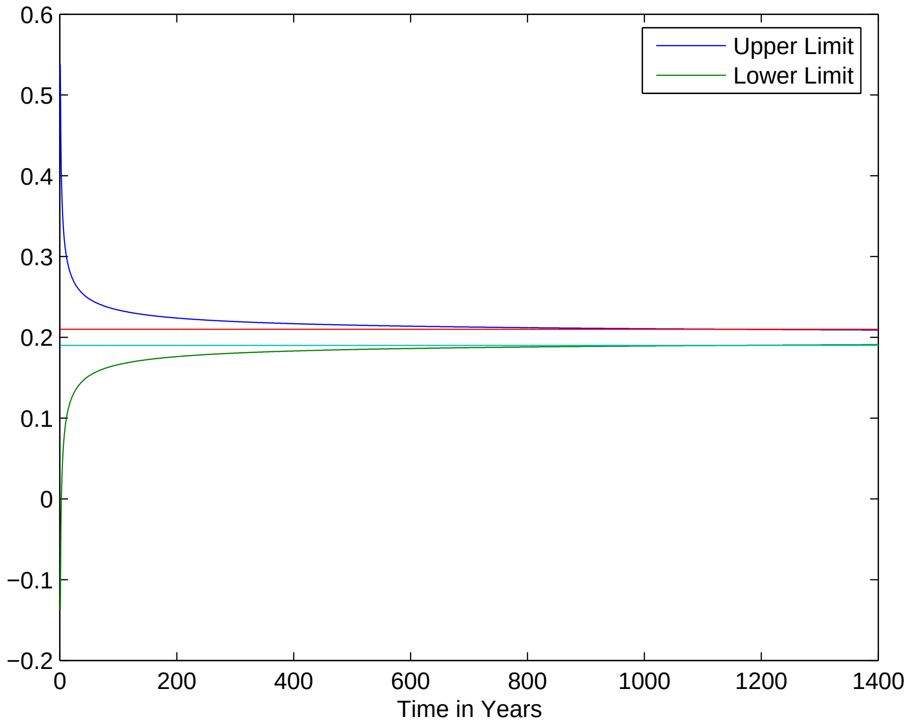

Figure 5.1: Variation of confidence interval with number of years. It takes many years of data to obtain a reasonable estimate of mean of a stock.

n periods. The estimate itself is normally distributed with mean μ and variance σ^2/n. Hence the implemented weight in risky asset is $\hat{\pi} = \frac{1}{\gamma\sigma^2}\hat{\mu}$. The expected mean variance utility as a function of μ is:

$$\begin{aligned}\hat{U}(\mu) &= \mathbb{E}_\mu\left[\left(\frac{1}{\gamma\sigma^2}\hat{\mu}\right)\mu - \frac{\gamma}{2}\left(\frac{1}{\gamma\sigma^2}\hat{\mu}\right)\sigma^2\left(\frac{1}{\gamma\sigma^2}\hat{\mu}\right)\right] \\ &= \frac{1}{2\gamma}\left(\frac{\mu^2}{\sigma^2} - \frac{1}{n}\right)\end{aligned} \tag{5.2}$$

For $\mu \approx \sigma$, the percentage loss due to uncertainty in return vector is $\frac{100}{n}\%$. This loss can be substantial for small values of n.

We can generalize the above to m stocks. Suppose we have a reasonably accurate estimate of Σ but an estimate of return is obtained from the past data. In general covariance matrix can be estimated more precisely by using high frequency data. In addition, the effect of uncertainty in covariance is usually smaller as compared to the uncertainty in returns [CZ93]. Suppose $\hat{\boldsymbol{\mu}}$ is an estimator of $\boldsymbol{\mu}$ with $\mathbb{E}[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$ and $Cov(\hat{\boldsymbol{\mu}}) = \mathcal{B}$. If $\hat{\boldsymbol{\mu}}$ is the maximum likelihood estimate of $\boldsymbol{\mu}$ based on past n periods of data, then $Cov(\hat{\boldsymbol{\mu}}) = \Sigma/n$. The expected utility as a function of $\boldsymbol{\mu}$ when portfolio weights $\hat{\boldsymbol{\pi}} = \frac{1}{\gamma}\Sigma^{-1}\hat{\boldsymbol{\mu}}$ is:

$$\begin{aligned}\hat{U}(\boldsymbol{\mu}) &= \mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\pi}}'\boldsymbol{\mu} - \frac{1}{2}\gamma\hat{\boldsymbol{\pi}}'\Sigma\hat{\boldsymbol{\pi}}\right] \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\left(\frac{1}{\gamma}\Sigma^{-1}\hat{\boldsymbol{\mu}}\right)'\boldsymbol{\mu} - \frac{1}{2}\gamma\left(\frac{1}{\gamma}\Sigma^{-1}\hat{\boldsymbol{\mu}}\right)'\Sigma\left(\frac{1}{\gamma}\Sigma^{-1}\hat{\boldsymbol{\mu}}\right)\right] \\ &= \frac{1}{\gamma}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} + \mathrm{tr}\left(\mathcal{B}\Sigma^{-1}\right)\right)\right) \\ &= \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \mathrm{tr}\left(\mathcal{B}\Sigma^{-1}\right)\right).\end{aligned} \tag{5.3}$$

If the covariance matrix $\mathcal{B}$ is Σ/n, as in the case of maximum likelihood estimate based on n period data, then the expected utility is:

$$\hat{U}(\boldsymbol{\mu}) = \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{m}{n}\right) \tag{5.4}$$

As we can see, if the number of stocks m is large and n is comparatively small, the loss in utility function due to uncertainty can be large.

5.2 Robust Optimization Approach

Robust optimization is one way to account for the error in estimating parameter $\boldsymbol{\mu}$. Suppose it is known or strongly believed that the parameter lies in a convex uncertainty set $\mathcal{U}$. A min-max based robust optimization approach would solve the following problem:

$$\max_{\boldsymbol{\pi}} \min_{\mu \in \mathcal{U}} \boldsymbol{\pi}' \boldsymbol{\mu} - \frac{1}{2} \gamma \boldsymbol{\pi} \Sigma \boldsymbol{\pi}. \tag{5.5}$$

Because the set $\mathcal{U}$ is convex and the objective is linear in $\boldsymbol{\mu}$, we can exchange the order of min-max in (5.5). Therefore Formulation (5.5) is equivalent to:

$$\min_{\mu \in \mathcal{U}} \max_{\boldsymbol{\pi}} \boldsymbol{\pi}' \boldsymbol{\mu} - \frac{1}{2} \gamma \boldsymbol{\pi} \Sigma \boldsymbol{\pi}. \tag{5.6}$$

Solving the inner optimization problem with respect to $\boldsymbol{\pi}$, we obtain

$$\frac{1}{2\gamma} \min_{\mu \in \mathcal{U}} \left(\boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right). \tag{5.7}$$

As can be seen, if we use robust optimization we obtain the worst possible utility from the chosen uncertainty set. One may argue that such a conservative approach is desirable in a highly volatile market; however, such a conservative approach may not guarantee a better solution than classical estimation based policy. The operational statistics approach defined in the next section strives to build a policy better than classical estimation based policy.

5.3 Operational Statistics Approach to Portfolio Optimization

Operational statistics is defined in Chapter 4. First, we need to choose a function class, i.e. a set of functions which define mapping of past data to a policy. We can restrict our attention to functions of $\hat{\boldsymbol{\mu}}$ (the maximum likelihood estimate of $\boldsymbol{\mu}$) as $\hat{\boldsymbol{\mu}}$ is a sufficient statistics for $\boldsymbol{\mu}$. Suppose the function class we choose is $\mathcal{F}$ and the corresponding policy set is $\Pi_{\mathcal{F}} = \{\pi : \pi = f(\hat{\boldsymbol{\mu}}), f \in \mathcal{F}\}$. We choose the function class $\mathcal{F}$ in such a way that $\hat{\boldsymbol{\pi}} \in \Pi_{\mathcal{F}}$. This will guarantee the feasibility of the problem we want to solve. In addition, suppose $\tilde{\boldsymbol{\mu}}$ is a representative return value that is subjectively chosen by the decision maker. The subjectivity in choosing $\tilde{\boldsymbol{\mu}}$ would allow expert opinion to be incorporated in the optimization problem. However, $\tilde{\boldsymbol{\mu}}$ can also be chosen objectively, for example, using maximum likelihood estimate $\hat{\boldsymbol{\mu}}$.

The operational statistics optimization problem is as follows:

$$\begin{aligned}
&\max_f \mathbb{E}_{\tilde{\boldsymbol{\mu}}}\left[f(\hat{\boldsymbol{\mu}})'\tilde{\boldsymbol{\mu}} - \frac{1}{2}f(\hat{\boldsymbol{\mu}})'\Sigma f(\hat{\boldsymbol{\mu}})\right]\\
&\text{subject to:}\\
&\mathbb{E}_{\boldsymbol{\mu}}\left[f(\hat{\boldsymbol{\mu}})'\boldsymbol{\mu} - \frac{1}{2}f(\hat{\boldsymbol{\mu}})'\Sigma f(\hat{\boldsymbol{\mu}})\right] \geq \hat{U}(\boldsymbol{\mu}), \qquad \forall \boldsymbol{\mu} \in \mathcal{U}.
\end{aligned} \tag{5.8}$$

The function class we choose to solve Formulation (5.8) is $\mathcal{F} = \{\frac{1}{\gamma}\Sigma^{-1}A\hat{\boldsymbol{\mu}} : \Sigma^{-1}A \in S^n, A \in S^n\}$. The set S^n is the set of all symmetric $n \times n$ matrices. Later in Section 5.5 we connect our choice of the function set with norm constrained portfolio. We choose the uncertainty set $\mathcal{U}$ to be an ellipsoid centered around $\tilde{\boldsymbol{\mu}}$. Note that the usual confidence region for multivariate normal is an ellipsoid centered around $\hat{\boldsymbol{\mu}}$, i.e $\mathcal{U} = \{\boldsymbol{x} : (\boldsymbol{x}-\tilde{\boldsymbol{\mu}})^T G(\boldsymbol{x}-\tilde{\boldsymbol{\mu}}) \leq 1\}$ where G is a positive semi-definite matrix. The assumption that $\tilde{\boldsymbol{\mu}}$ is the center of the ellipsoid $\mathcal{U}$ is made for simplicity of notation and can be relaxed.

The problem we now have is:

Portfolio OS Problem 1

$$\begin{aligned}
&\max_A \mathbb{E}_{\tilde{\boldsymbol{\mu}}}\left[\hat{\boldsymbol{\mu}}'A\Sigma^{-1}\tilde{\boldsymbol{\mu}} - \frac{1}{2}\hat{\boldsymbol{\mu}}'A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right]\\
&\text{subject to: } \mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}'A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}'A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] \geq \frac{1}{2}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{m}{n}\right), \qquad \forall \boldsymbol{\mu} \in \mathcal{U}.
\end{aligned} \tag{5.9}$$

The objective function in Formulation (5.9) is concave in matrix A. We also have infinitely many convex constraints. It is not clear whether the problem can be solved efficiently. Next we will show that the problem can be reformulated as a semidefinite program (SDP). An SDP is a generalization of linear program which concerns with optimization of a linear function over intersection of a cone of semidefinite matrices and an affine space. An SDP can be solved efficiently by using interior point methods. In its dual form a semidefinite program can be expressed as:

$$\begin{aligned}
&\max_{\boldsymbol{y}, S_1, \ldots, S_n} \boldsymbol{b}'\boldsymbol{y}\\
&\text{subject to:}\\
&\qquad y_1 A_{11} + \cdots + y_m A_{m1} + S_1 = C_1\\
&\qquad \cdots\\
&\qquad y_1 A_{1n} + \cdots + y_m A_{mn} + S_n = C_n\\
&\qquad S_i \succeq 0 \qquad \text{for } i = 1 \text{ to } n,
\end{aligned} \tag{5.10}$$

where A's and C's are $n \times n$ matrices and are parameters of the system. The constraint $S_i \succeq 0$ means that the matrix S_i is a symmetric positive semidefinite matrix. Note that a linear constraint $A\boldsymbol{y} = c$ or a second order conic constraint $\|\boldsymbol{y}\|_2 \leq z$ can be expressed as semidefinite constraints. For more on semidefinite programs and second order cone programs see [LS91, VB93, NN94, BTN01]. We now have the following theorem:

Theorem 10. *The Problem* (5.9) *can be reformulated as a semidefinite program.*

Proof. To prove the theorem we need the following two lemmas:

Lemma 11. ***(S-lemma)*** *Let B_1 and B_2 be symmetric matrices and let $H_i(x) = \boldsymbol{x}'B_i\boldsymbol{x} + 2\boldsymbol{c_i'}\boldsymbol{x} + \boldsymbol{d_i}, i = 0, 1$ be two quadratic functions of $\boldsymbol{x} \in \mathbb{R}^n$. If there exists a $\boldsymbol{y}$ such that $H_1(\boldsymbol{y}) > 0$, then $H_0(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$ satisfying $H_1(\boldsymbol{x}) \geq 0$, if and only if there exists a $\lambda \geq 0$ such that*

$$\begin{bmatrix} d_0 & c_0' \\ c_0 & B_0 \end{bmatrix} - \lambda \begin{bmatrix} d_1 & c_1' \\ c_1 & B_1 \end{bmatrix} \succeq 0$$

Lemma 12. ***(Schur Complement)*** *Let*

$$W = \begin{bmatrix} X & Y' \\ Y & Z \end{bmatrix}$$

be a symmetric matrix with $p \times p$ block X and $q \times q$ block Z. Assume that X is positive definite. Then W is positive semi-definite if and only if the Schur complement of W in X, i.e., the matrix

$$Z - YX^{-1}Y'$$

is positive semi-definite.

First, note that

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}'A\Sigma^{-1}\boldsymbol{\mu}\right] = \boldsymbol{\mu}'A\Sigma^{-1}\boldsymbol{\mu} \tag{5.11}$$

and

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}'A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathrm{tr}\left(\hat{\boldsymbol{\mu}}'A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right)\right] \\
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathrm{tr}\left(\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}'A\Sigma^{-1}A\right)\right] \\
&= \mathrm{tr}\left(\mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}'\right]A\Sigma^{-1}A\right) \\
&= \mathrm{tr}\left(\left(\boldsymbol{\mu}\boldsymbol{\mu}' + \frac{1}{n}\Sigma\right)A\Sigma^{-1}A\right) \\
&= \mathrm{tr}\left(\boldsymbol{\mu}\boldsymbol{\mu}'A\Sigma^{-1}A\right) + \frac{1}{n}\mathrm{tr}\left(\Sigma^{1/2}\Sigma^{1/2}A\Sigma^{-1/2}\Sigma^{-1/2}A\right) \\
&= \boldsymbol{\mu}'A\Sigma^{-1}A\boldsymbol{\mu} + \frac{1}{n}\mathrm{tr}\left(\Sigma^{1/2}A\Sigma^{-1/2}\Sigma^{-1/2}A\Sigma^{1/2}\right) \\
&= \boldsymbol{\mu}'A\Sigma^{-1}A\boldsymbol{\mu} + \frac{1}{n}\mathrm{tr}\left(X'X\right),
\end{aligned} \tag{5.12}$$

where $X = \Sigma^{-1/2} A \Sigma^{1/2}$.

The constraints can be rewritten as:

$$\left(\boldsymbol{\mu}' A \Sigma^{-1} A \boldsymbol{\mu} + \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} - 2\boldsymbol{\mu}' A \Sigma^{-1} \boldsymbol{\mu}\right) \leq \frac{1}{n}(m - \mathrm{tr}(X'X)) \tag{5.13}$$

$$\Rightarrow \left(\boldsymbol{\mu}'(A - I)\Sigma^{-1}(A - I)\boldsymbol{\mu}\right) \leq \frac{1}{n}(m - \mathrm{tr}(X'X)). \tag{5.14}$$

Note that the right hand side is independent of $\boldsymbol{\mu}$ in the above inequality. We can now write the constraints set as follows:

$$\left(\boldsymbol{\mu}'(A - I)\Sigma^{-1}(A - I)\boldsymbol{\mu}\right) \leq \nu \qquad \forall \boldsymbol{\mu} \in \mathcal{U} \tag{5.15}$$

$$\nu \leq \frac{1}{n}(m - \mathrm{tr}(X'X)). \tag{5.16}$$

Let $X_i, i = 1, \ldots, m$ be the column vectors of matrix X. Constraint (5.16) is a second order conic constraint since it can be written as:

$$\mathrm{tr}(X'X) \leq m - n\nu \tag{5.17}$$

$$\Rightarrow \sum_{i=1}^{l} X_i' X_i \leq \frac{(m - n\nu + 1)^2}{4} - \frac{(m - n\nu - 1)^2}{4} \tag{5.18}$$

$$\Rightarrow \left\| \begin{pmatrix} X_1 \\ \vdots \\ X_m \\ \frac{m-1-n\nu}{2} \end{pmatrix} \right\|_2 \leq \frac{m + 1 - n\nu}{2}. \tag{5.19}$$

Constraint (5.15) is equivalent to

$$\nu - \boldsymbol{\mu}'(A - I)\Sigma^{-1}(A - I)\boldsymbol{\mu} \geq 0$$

for all b such that $1 - (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}})' G (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) \geq 0$. To proceed further, we need the following lemma:

Note that for $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$, the inequality $1 - (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}})' G (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) \geq 0$ is strictly feasible. Hence Constraint (5.15) can be rewritten as:

$$\begin{bmatrix} \nu - \lambda(1 - \tilde{\boldsymbol{\mu}} G \tilde{\boldsymbol{\mu}}) & -\lambda \tilde{\boldsymbol{\mu}}' G \\ -\lambda G \tilde{\boldsymbol{\mu}} & \lambda G - (A - I)\Sigma^{-1}(A - I) \end{bmatrix} \succeq 0 \tag{5.20}$$

Consider the following two cases:

1. $\tilde{\boldsymbol{\mu}} = 0$. Constraint (5.20) in this case is:

$$\begin{bmatrix} \nu - \lambda & 0 \\ 0 & \lambda G - (A-I)\Sigma^{-1}(A-I) \end{bmatrix} \succeq 0 \tag{5.21}$$

which translates to the following two constraints:

$$\nu \geq \lambda; \quad \lambda G - (A-I)\Sigma^{-1}(A-I) \succeq 0 \tag{5.22}$$

Since Σ is positive definite, we can write Constraint (5.20) using Schur Complement lemma as:

$$\nu \geq \lambda; \quad \begin{bmatrix} \Sigma & A-I \\ A-I & \lambda G \end{bmatrix} \succeq 0 \tag{5.23}$$

The above is a positive semi-definite constraint set.

2. $\tilde{\boldsymbol{\mu}} \neq 0$. First, note that if $\nu - \lambda(1 - \tilde{\boldsymbol{\mu}} G \tilde{\boldsymbol{\mu}}) = 0$ then for the matrix in Constraint (5.20) to be positive semi-definite λ should be equal to zero. This in turn implies $\nu = 0$, and therefore $A = I$. For any non-trivial solution $\nu - \lambda(1 - \tilde{\boldsymbol{\mu}} G \tilde{\boldsymbol{\mu}}) > 0$ should be true. Applying the Schur Complement lemma we obtain:

$$\begin{bmatrix} \nu - \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}}) & -\lambda \tilde{\boldsymbol{\mu}}' G \\ -\lambda G \tilde{\boldsymbol{\mu}} & \lambda G - (A-I)\Sigma^{-1}(A-I) \end{bmatrix} \succeq 0 \tag{5.24}$$

$$\iff \lambda G - (A-I)\Sigma^{-1}(A-I) - \frac{\lambda^2}{\nu - \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})} G \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}' G \succeq 0. \tag{5.25}$$

Constraint (5.25) can also be rewritten as:

$$\lambda G - (A-I)\Sigma^{-1}(A-I) - \alpha G \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}' G \succeq 0 \tag{5.26}$$

$$\frac{\lambda^2}{\nu - \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})} \leq \alpha. \tag{5.27}$$

Constraint (5.26) can be written using Schur Complement lemma as:

$$\begin{bmatrix} \Sigma & A-I \\ A-I & \lambda G - \alpha G \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}' G \end{bmatrix} \succeq 0, \tag{5.28}$$

which is a positive semi-definite constraint. Constraint (5.27) can be expressed in conic form as:

$$\left\| \begin{pmatrix} \lambda \\ \frac{\alpha - \nu + \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2} \end{pmatrix} \right\|_2 \leq \frac{\alpha + \nu - \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2} \tag{5.29}$$

The objective in Formulation (5.9) is

$$\max_A \tilde{\boldsymbol{\mu}}' A \Sigma^{-1} \tilde{\boldsymbol{\mu}} - \frac{1}{2}\left(\tilde{\boldsymbol{\mu}}' A \Sigma^{-1} A \tilde{\boldsymbol{\mu}} + \frac{1}{n}\text{tr}(X'X) \right) \tag{5.30}$$

Completing the square in Equation (5.30) we obtain

$$\max_A -\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}'(A-I)\Sigma^{-1}(A-I)\tilde{\boldsymbol{\mu}} + \frac{1}{n}\text{tr}(X'X) \right) + \frac{1}{2}\tilde{\boldsymbol{\mu}}'\Sigma^{-1}\tilde{\boldsymbol{\mu}} \tag{5.31}$$

As $\frac{1}{2}\tilde{\boldsymbol{\mu}}'\Sigma^{-1}\tilde{\boldsymbol{\mu}}$ is a constant, the objective is equivalent to minimizing:

$$\min_A \left(n\tilde{\boldsymbol{\mu}}'(A-I)\Sigma^{-1}(A-I)\tilde{\boldsymbol{\mu}} + \text{tr}(X'X) \right) \tag{5.32}$$

The objective can therefore be written as:

$$\min z$$

subject to:

$$\left(n\tilde{\boldsymbol{\mu}}'(A-I)\Sigma^{-1}(A-I)\tilde{\boldsymbol{\mu}} + \text{tr}(X'X) \right) \leq z \tag{5.33}$$

Let $y = n^{1/2}\Sigma^{-1/2}A\tilde{\boldsymbol{\mu}}$. We can write (5.33) as a conic constraint:

$$\begin{aligned} \left\| \begin{pmatrix} y \\ X_1 \\ \vdots \\ X_m \end{pmatrix} \right\|_2 &\leq z \\ y &= n^{1/2}\Sigma^{-1/2}(A-I)\tilde{\boldsymbol{\mu}} \end{aligned} \tag{5.34}$$

The formulation (5.9) is now the following semi-definite program:

$$\min z$$

subject to:

$$
\begin{aligned}
\left\| \begin{pmatrix} y \\ X_1 \\ \vdots \\ X_m \end{pmatrix} \right\|_2 &\le z \\
\left\| \begin{pmatrix} X_1 \\ \vdots \\ X_m \\ \frac{m-1-n\nu}{2} \end{pmatrix} \right\|_2 &\le \frac{m+1-n\nu}{2}, \\
\begin{bmatrix} \Sigma & A-I \\ A-I & \lambda G - \alpha G \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}' G \end{bmatrix} &\succeq 0, \\
\left\| \begin{pmatrix} \lambda \\ \frac{\alpha - \nu + \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2} \end{pmatrix} \right\|_2 &\le \frac{\alpha + \nu - \lambda(1 - \tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2}, \\
y &= n^{1/2} \Sigma^{-1/2} (A - I) \tilde{\boldsymbol{\mu}}, \\
\lambda &\ge 0
\end{aligned}
\tag{5.35}
$$

for $\tilde{\boldsymbol{\mu}} \neq 0$. The formulation for $\tilde{\boldsymbol{\mu}} = 0$ is similar. □

5.4 Related Literature

There is a long history of work on mean-variance and other related portfolio optimization problems. The foundation of modern portfolio theory is laid by the seminal work of Harry Markowitz [Mar52, Mar59] on mean variance portfolio selection. The beauty of Markowitz portfolio lies in its simplicity and numerical tractability. However, it is also well known that a mean variance portfolio thus constructed has poor out of sample performance due to large error in estimates, particularly of mean vector, but also of covariance matrix. One of the earlier papers that identifies the poor performance of Markowitz portfolio due to estimation error was by Jobson and Korkie [JK81]. Some other studies on the effect of estimation error on Markowitz portfolio are [FS88, Mic89, CZ93] and [Bro93]. It has also been observed that a naive Markowitz portfolio rarely outperforms a portfolio that assigns equal weights to each of its constituents [DGU09]. Over the years many remedies have been proposed by various authors to mitigate the effect of estimation error on the portfolio.

A Bayesian approach suggests to use a prior on mean or/and covariance matrix to reduce the uncertainty in parameters. A good choice of prior may reduce the effects of uncertainty considerably whereas a bad choice may make it even worse. In literature various choices of priors have been proposed. A popular choice is to use non informative diffuse priors [BBK79, Bar74]. A diffuse prior for unknown mean and known variance matrix is normally distributed with mean $\hat{\boldsymbol{\mu}}$ and variance $\Sigma\left(1+\frac{1}{n}\right)$ where n is the length of the period used to estimate $\boldsymbol{\mu}$. Another choice is using Bayes-Stein prior [Ste55, Jor86, JS61] which shrink the sample means $\mu_1, \ldots, \mu_m$ towards a common value. A Bayes-Stein estimator of the mean vector is:

$$(1-\hat{w})\hat{\boldsymbol{\mu}} + \hat{w}\bar{\mu}\boldsymbol{e} \tag{5.36}$$

where

$$\hat{w} = \min\left(1, \frac{(m-2)/n}{(\hat{\boldsymbol{\mu}} - \bar{\mu}\boldsymbol{e})\Sigma^{-1}(\hat{\boldsymbol{\mu}} - \bar{\mu}\boldsymbol{e})}\right). \tag{5.37}$$

The estimator shrinks the means to a common value $\bar{\mu}$.

Robust optimization methods in portfolio optimization consider the parameters of the problem to lie in an uncertainty set and use objectives like max-min or min-max regret to calculate robust portfolios. Goldfarb and Iyengar [GI03] considered the worst case optimization problem under interval and ellipsoidal uncertainties for mean vector and covariance matrix respectively and showed that the problem can be reformulated as a second order conic program (SOCP). In addition, see [UW03, TK04] for related research on worst case portfolio optimization.

It has also been observed that adding constraints have a positive shrinkage effect on the estimates and thus result in better portfolios. Examples of research along this line are: Constraining portfolios by putting short selling constraints [FS88, Cho93, JM03]; explicitly constraining the moments of a portfolio [MP00]; shrinking covariance matrix [LW04a, LW04b]; and constraining a vector norm of portfolio weights [DGNU09].

5.5 Connection to Norm-Constrained Portfolio

As mentioned in Section 5.4, putting constraints on mean variance portfolio optimization problem improves the performance of the portfolio in practice. One particular approach that has been explored by [DGNU09] is constraining the norm of portfolio weights to be less than some specified value. Various norms such as 1-norm, 2-norm and matrix-norm have been explored by [DGNU09]. We are particularly interested in the portfolio subject to matrix norm:

$$\begin{aligned} &\max_{\boldsymbol{\pi}} \ \boldsymbol{\pi}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{\pi}\Sigma\boldsymbol{\pi} \\ &\text{subject to} \\ &\boldsymbol{\pi}'X'\boldsymbol{\pi} \leq \mathbf{1}, \end{aligned} \tag{5.38}$$

An alternative characterization of (5.38) using penalty function is:

$$\max_{\boldsymbol{\pi}} \ \boldsymbol{\pi}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{\pi}\Sigma\boldsymbol{\pi} - \boldsymbol{\pi}'X'\boldsymbol{\pi}. \tag{5.39}$$

Formulation (5.39) has the solution

$$\boldsymbol{\pi}_{norm} = \frac{1}{\gamma}\Sigma^{-1}(I + \gamma\Sigma X)\boldsymbol{\mu}. \tag{5.40}$$

The implemented policy would be

$$\hat{\boldsymbol{\pi}}_{norm} = \frac{1}{\gamma}\Sigma^{-1}(I + \gamma\Sigma X)\hat{\boldsymbol{\mu}}, \tag{5.41}$$

where I is an identity matrix of suitable dimension. The operational statistics policy $\boldsymbol{\pi}_{os}$ is

$$\boldsymbol{\pi}_{os} = \frac{1}{\gamma}\Sigma^{-1}A^*\hat{\boldsymbol{\mu}}, \tag{5.42}$$

where A^* is the solution of the operational statistics optimization problem.

Comparing (5.41) and (5.42) we see a parallel between these two solutions. If $A^* = I + \gamma\Sigma X$ then both the solutions are same. Hence the operational statistics solution is the solution of a norm constrained portfolio problem with matrix X suitably defined. One difference is that the matrix in operational statistics X would be chosen as a solution of an optimization problem.

5.6 Extensions

In this section we consider various extensions to our base operational statistics model.

5.6.1 Using a Prior

One of the difficulty in implementing the operational statistics portfolio optimization problem is how to choose $\tilde{\boldsymbol{\mu}}$ subjectively. In practice, in the absence of any subjective information one may be inclined to choose the maximum likelihood estimate $\hat{\boldsymbol{\mu}}$ as $\tilde{\boldsymbol{\mu}}$. This may not be desirable as we already know that $\hat{\boldsymbol{\mu}}$ is a poor proxy for $\boldsymbol{\mu}$. In this case a good choice is to use a prior distribution on set $\mathcal{U}$ similar to the operational statistics Formulation (4.13) in Chapter 4. Let $P_{\boldsymbol{\mu}}$ be our choice of prior on set $\mathcal{U}$.

$$\begin{aligned} &\max_A \int_{\mathcal{U}} \mathbb{E}_{\mu}\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] dP_{\boldsymbol{\mu}} \\ &\text{subject to: } \mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] \geq \frac{1}{2}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{m}{n}\right), \qquad \forall \boldsymbol{\mu} \in \mathcal{U}. \end{aligned} \tag{5.43}$$

One particular choice of prior we consider in our numerical experiments is a uniform prior on set $\mathcal{U}$. This also corresponds to non-informative prior in the absence of any subjective information. For uniform prior we have the following problem:

Portfolio OS Problem 2

$$\begin{aligned}
&\max_A \frac{1}{|\mathcal{U}|}\int_{\mathcal{U}} \mathbb{E}_\mu\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] d\boldsymbol{\mu} \\
&\quad \text{subject to: } \mathbb{E}_{\boldsymbol{\mu}}\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] \geq \frac{1}{2}\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{m}{n}\right), \qquad \forall \boldsymbol{\mu} \in \mathcal{U}.
\end{aligned} \tag{5.44}$$

Proposition 13. *The Problem* (5.44) *can be reformulated as a semidefinite program.*

Proof. Note that the constraint set is same as (5.9). So we just need to focus on the objective. From (5.12)

$$\begin{aligned}
\mathbb{E}_\mu\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] &= \boldsymbol{\mu}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}' A\Sigma^{-1}A\boldsymbol{\mu} - \frac{1}{2n}\mathrm{tr}(X'X) \\
&= \boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\boldsymbol{\mu} - \frac{1}{2n}\mathrm{tr}(X'X)
\end{aligned} \tag{5.45}$$

Hence

$$\int_{\mathcal{U}} \mathbb{E}_\mu\left[\hat{\boldsymbol{\mu}}' A\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\hat{\boldsymbol{\mu}}' A\Sigma^{-1}A\hat{\boldsymbol{\mu}}\right] d\boldsymbol{\mu} = \int_{\mathcal{U}}\left[\boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\boldsymbol{\mu}\right] d\boldsymbol{\mu} - \frac{|\mathcal{U}|}{2n}\mathrm{tr}(X'X) \tag{5.46}$$

Now

$$\begin{aligned}
\int_{\mathcal{U}}\left[\boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\boldsymbol{\mu}\right] d\boldsymbol{\mu} &= \int_{\mathcal{U}} \mathrm{tr}\left(\boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\boldsymbol{\mu}\right) d\boldsymbol{\mu} \\
&= \int_{\mathcal{U}} \mathrm{tr}\left(\boldsymbol{\mu}\boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\right) d\boldsymbol{\mu} \\
&= \mathrm{tr}\left(\left(\int_{\mathcal{U}} \boldsymbol{\mu}\boldsymbol{\mu}' d\boldsymbol{\mu}\right)\ (A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\right)
\end{aligned} \tag{5.47}$$

Recall that the set $\mathcal{U} = \{\boldsymbol{x} : (\boldsymbol{x} - \tilde{\boldsymbol{\mu}})^T G(\boldsymbol{x} - \tilde{\boldsymbol{\mu}}) \leq 1\}$. Let L be the symmetric square root of positive definite matrix G. By substituting $\boldsymbol{z} = L(\boldsymbol{x} - \tilde{\boldsymbol{\mu}})$ we obtain

$$\int_{\mathcal{U}} \boldsymbol{\mu}\boldsymbol{\mu}' d\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}'|\mathcal{U}| + \det(L^{-1})|\int_{\{\boldsymbol{z}'\boldsymbol{z}\leq 1\}} L^{-1}\boldsymbol{z}\boldsymbol{z}'L^{-1} d\boldsymbol{z} \tag{5.48}$$

Note that if $\boldsymbol{z} = \{z_1, \ldots, z_i, \ldots, z_n\}$ then

$$\int_{\{\boldsymbol{z}'\boldsymbol{z}\leq 1\}} z_i^2 d\boldsymbol{z} = C_n \frac{1}{n+2} \tag{5.49}$$

and for $i \neq j$

$$\int_{\{\boldsymbol{z}'\boldsymbol{z}\le 1\}} z_i z_j d\boldsymbol{z} = 0. \tag{5.50}$$

where C_n is the volume of $n-$dimensional unit sphere. In addition, note that $|\mathcal{U}| = C_n|\det(L^{-1})|$. Therefore,

$$\int_{\mathcal{U}} \boldsymbol{\mu}\boldsymbol{\mu}' d\boldsymbol{\mu} = |\mathcal{U}| \left(\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}' + \frac{1}{n+2}L^{-1}L^{-1}\right) \tag{5.51}$$

and

$$\begin{aligned}\int_{\mathcal{U}} \left[\boldsymbol{\mu}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\boldsymbol{\mu}\right] d\boldsymbol{\mu} &= |\mathcal{U}|\mathrm{tr}\left(\left(\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}' + \frac{1}{n+2}L^{-1}L^{-1}\right)(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\right)\\ &= |\mathcal{U}|\mathrm{tr}\left(\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\right)\\ &\quad + \frac{|\mathcal{U}|}{n+2}\mathrm{tr}\left(L^{-1}L^{-1}(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\right)\\ &= |\mathcal{U}|\left(\tilde{\boldsymbol{\mu}}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\tilde{\boldsymbol{\mu}}\right)\\ &\quad + \frac{|\mathcal{U}|}{n+2}\mathrm{tr}\left(L^{-1}(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)L^{-1}\right)\end{aligned} \tag{5.52}$$

Hence the objective function of (5.44) is

$$\tilde{\boldsymbol{\mu}}'(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)\tilde{\boldsymbol{\mu}} - \frac{1}{2n}\mathrm{tr}(X'X) + \frac{1}{n+2}\mathrm{tr}\left(L^{-1}(A\Sigma^{-1} - \frac{1}{2}A\Sigma^{-1}A)L^{-1}\right) \tag{5.53}$$

Completing the squares in (5.53) we obtain

$$\begin{aligned}-\frac{1}{2}\left(\tilde{\boldsymbol{\mu}}'(A-I)\Sigma^{-1}(A-I)\tilde{\boldsymbol{\mu}}\right) - \frac{1}{2n}\mathrm{tr}(X'X) - \frac{1}{2(n+2)}\mathrm{tr}\left(L^{-1}(A-I)\Sigma^{-1}(A-I)L^{-1}\right)\\ + \frac{1}{2}\tilde{\boldsymbol{\mu}}'\Sigma^{-1}\tilde{\boldsymbol{\mu}} + \frac{1}{2(n+2)}\mathrm{tr}\left(L^{-1}\Sigma^{-1}(A-I)\right)\end{aligned} \tag{5.54}$$

The last two terms in (5.54) are constants and can be dropped. Hence the objective is equivalent to minimizing:

$$n\tilde{\boldsymbol{\mu}}'(A-I)\Sigma^{-1}(A-I)\tilde{\boldsymbol{\mu}} + \mathrm{tr}(X'X) + \frac{n}{n+2}\mathrm{tr}\left(L^{-1}(A-I)\Sigma^{-1}(A-I)L^{-1}\right) \tag{5.55}$$

Let $W = \left(\frac{n}{n+2}\right)^{1/2} \Sigma^{-1/2}(A-I)L^{-1}$. Then, similar to the proof of Theorem 10, Formulation (5.44) can be written as:

$$\min z$$

subject to:

$$\begin{aligned}
\left\| \begin{pmatrix} y \\ X_1 \\ \vdots \\ X_m \\ W_1 \\ \vdots \\ W_m \end{pmatrix} \right\|_2 &\leq z \\
\left\| \begin{pmatrix} X_1 \\ \vdots \\ X_m \\ \frac{m-1-n\nu}{2} \end{pmatrix} \right\|_2 &\leq \frac{m+1-n\nu}{2}, \\
\begin{bmatrix} \Sigma & A-I \\ A-I & \lambda G - \alpha G \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}' G \end{bmatrix} &\succeq 0, \\
\left\| \begin{pmatrix} \lambda \\ \frac{\alpha-\nu+\lambda(1-\tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2} \end{pmatrix} \right\|_2 &\leq \frac{\alpha+\nu-\lambda(1-\tilde{\boldsymbol{\mu}}' G \tilde{\boldsymbol{\mu}})}{2}, \\
y &= n^{1/2}\Sigma^{-1/2}(A-I)\tilde{\boldsymbol{\mu}}, \\
\lambda &\geq 0
\end{aligned} \tag{5.56}$$

□

5.6.2 Portfolio with No Risk-Free Asset

We can also consider a portfolio with no risk-free asset. Here the vector $\boldsymbol{\mu}$ represents true returns, not the excess returns. In addition, the weights of the constituents should sum to one. We have the following nominal problem:

$$\begin{gathered}
\max_{\boldsymbol{\pi}} \boldsymbol{\pi}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{\pi}\Sigma\boldsymbol{\pi} \\
\text{subject to} \\
\boldsymbol{\pi}'\boldsymbol{e} = 1
\end{gathered} \tag{5.57}$$

The solution of 5.57 is given by

$$\boldsymbol{\pi} = \frac{1}{\gamma}\Sigma^{-1}\left(\boldsymbol{\mu} - \mu_0\boldsymbol{e}\right) \tag{5.58}$$

where

$$\mu_0 = \frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{e} - \gamma}{\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e}}. \tag{5.59}$$

The classical estimation based policy would replace $\boldsymbol{\mu}$ by its point estimate $\hat{\boldsymbol{\mu}}$. The corresponding operational statistics problem is:

Portfolio OS Problem 3

$$\max_A \mathbb{E}_{\tilde{\mu}}\left[(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})'\Sigma^{-1}\tilde{\boldsymbol{\mu}} - \frac{1}{2}(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})\Sigma^{-1}(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})\right]d\boldsymbol{\mu}$$

subject to:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\mu}}\left[(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})\Sigma^{-1}(\hat{\boldsymbol{\mu}}A - \hat{\mu}_1\boldsymbol{e})\right] \\
&\quad \geq \mathbb{E}_{\boldsymbol{\mu}}\left[(\hat{\boldsymbol{\mu}} - \hat{\mu}_0\boldsymbol{e})\Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}(\hat{\boldsymbol{\mu}} - \hat{\mu}_0\boldsymbol{e})\Sigma^{-1}(\hat{\boldsymbol{\mu}} - \hat{\mu}_0\boldsymbol{e})\right], \qquad \forall \boldsymbol{\mu} \in \mathcal{U},
\end{aligned} \tag{5.60}$$

where the policy space is parametrized by the positive semidefinite matrix A and

$$\hat{\mu}_1 = \frac{\hat{\boldsymbol{\mu}}'\Sigma^{-1}\boldsymbol{e} - \gamma}{\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e}}, \qquad \hat{\mu}_1 = \frac{\hat{\boldsymbol{\mu}}'A\Sigma^{-1}\boldsymbol{e} - \gamma}{\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e}}. \tag{5.61}$$

We state the next proposition without the proof, as the proof is similar to Proposition 13 and Theorem 10.

Proposition 14. *The Problem* (5.60) *can be reformulated as a semidefinite program.*

5.7 Numerical Results

For numerical experiments we chose a portfolio of 10 stocks. Each component of the return vector $\boldsymbol{u}$ is randomly generated between $[-0.05, 0.20]$. The covariance matrix Σ is randomly generated with an identity matrix (multiplied by a small constant) added to it so that Σ is well conditioned. The uncertainty set $\mathcal{U}$ is chosen to be:

$$\mathcal{U} = \{\boldsymbol{x} : (\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \leq \alpha\}.$$

The uncertainty set is centered around the true return vector μ and is parametrized by α. We compare the expected utilities of sample mean based policy, operational statistics policy (with uniform prior) and robust policy on 3 trial runs. All semidefinite programs are solved using SDPT3 semidefinite program solver using the MATLAB interface YALMIP.

Table 5.1 compares the expected utility of the three policies. The operational statistics policy with uniform prior is better than the sample mean based policy for all choices of α. It should be so because of the constraints in the operational statistics formulation. When α is very small or the uncertainty set is very small, the robust policy outperforms operational statistics policy. This is because when the size of the uncertainty set is very small, even the worst case model in the set is quite good. However, for large values of α, the operational statistics policy is better than the robust policy. When $\{\boldsymbol{\mu} = 0\} \in \mathcal{U}$, then the robust utility is equal to zero and the robust utility is even worse than the utility of sample mean based policy. Even for very large values of α, it is surprising that the operational statistics policy improves the sample mean based policy substantially. One caveat however is that the robust utility is lower bounded by zero whereas the operational statistics utility is lower bounded by the utility of sample mean policy. Therefore, when the utility of sample mean policy is negative as in experiment 3, the operational statistics utility also becomes negative for large values of α whereas the robust utility remains lower bounded by 0.

(Experiment 1: True Utility = 2.2375

α	Sample Mean Policy	Robust Policy	OS Policy
0.001	1.3328	**2.2312**	1.7824
0.01	1.3328	**2.1750**	1.7699
0.1	1.3328	1.6125	**1.6252**
1	1.3328	0	**1.5128**
10	1.3328	0	**1.4974**
100	1.3328	0	**1.4953**
1000	1.3328	0	**1.4951**
10000	1.3328	0	**1.4950**
100000	1.3328	0	**1.4950**

Experiment 2: True Utility = 1.2047

α	Sample Mean Policy	Robust Policy	OS Policy
0.001	0.4291	**1.1984**	0.8278
0.01	0.4291	**1.1422**	0.8150
0.1	0.4291	0.5797	**0.6980**
1	0.4291	0	**0.5981**
10	0.4291	0	**0.5793**
100	0.4291	0	**0.5772**
1000	0.4291	0	**0.5770**
10000	0.4291	0	**0.5770**
100000	0.4291	0	**0.5770**

Experiment 3: True Utility = 0.5226

α	Sample Mean Policy	Robust Policy	OS Policy
0.001	-0.1651	**0.4086**	0.2472
0.01	-0.1651	**0.3523**	0.2219
0.1	-0.1651	0	**0.0252**
1	-0.1651	0	**0.0047**
10	-0.1651	**0**	-0.1031
100	-0.1651	**0**	-0.1228
1000	-0.1651	**0**	-0.1249
10000	-0.1651	**0**	-0.1251
100000	-0.1651	**0**	-0.1251

Table 5.1: Comparison of utilities of operational statistics policy, worst case robust optimization policy and sample mean based policy

Chapter 6

Objective Operational Learning

Most of the research on stochastic modeling and optimization in Operations Research and Management Science (ORMS) is model-based. Standard model-based methods typically assume a parametric family of distributions and assume full knowledge of parameters. In practice, these parameters are estimated using data (e.g. Bayesian inference, point estimate, etc.), and then the model is optimized using estimated model. Though model-based methods remains the primary focus of ORMS literature, it is also well known that they can have poor out-of-sample performance and converge to unacceptable solutions if the underlying assumptions are incorrect (a dramatic illustration of this is the so-called spiral-down effect in revenue management; see [CdMK06]). On the other hand, while non-parametric or model free methods while guarantee asymptotic convergence to the optimal solution under minimal assumptions, they typically ignore structural or domain information about the problem, resulting in poor small sample performance. This limits the applicability of standard non-parametric model when systems are only locally stationary.

In this chapter we introduce the notion of *objective operational learning*. A distinguishing feature of this approach is that it incorporates structural knowledge and (possibly incorrect) assumptions about the problem but smoothly drops the incorrect ones and becomes increasingly non-parametric as the data size increases. The transition from an assumption-based estimate of the objective function to a non-parametric one is done with the aid of a kernel function. The objective operational learning approach aims to improve small sample performance while guaranteeing asymptotic convergence to the optimal solution under essentially the same conditions as other model-free non-parametric methods (even if the assumed structural model is incorrect).

In this chapter, we provide a general description of our algorithm and show how it can be used to learn the profit function and optimal order quantity for a newsvendor problem with a potentially order-dependent demand distribution. We make no assumptions about this dependence aside from continuity (under some metric) of demand distribution on order quantity. We provide a mechanism for choosing next period order quantity based on the current estimate of the profit function. The mechanism is a randomized decision rule which balances the classical trade off between exploration and exploitation.

The main contributions of our work can be summarized as follows:

(i) We introduce a new approach for learning the objective function and online optimization called objective operation learning. To our knowledge this is the first approach that incorporates possibly erroneous structural models in a systematic way to improve small sample performance while guaranteeing asymptotic convergence under minimal assumptions.

(ii) We show how domain information can be incorporated for different variations of the newsvendor problem. We show that the definition of our kernel function need not be similar to what is used in the nonparametric regression.

(iii) We apply our approach to newsvendor problem with order quantity dependent demand distribution. We prove that the point-wise convergence of estimates of profit function constructed using the operational statistics algorithm.

The rest of this chapter is organized as follows. In Section 6.1, we review the related literature. We introduce our problem setting in Section 6.2 and in Section 6.3, we briefly review non-paramteric regression. In Section 6.4, we study the newsvendor model in three scenarios and introduce the idea of objective operational statistics. Section 6.5 is the application of the objective operational learning method to newsvendor problem with order quantity dependent demand distribution and includes the convergence results.

6.1 Related Literature

Most of the non-parametric work in ORMS is without learning. Such studies find policies by bounding the worst case "loss" due to model error given a set of permissible models. Such a set is constructed using some nonparametric information about the underlying model. [Sca58, Gal92] and [MG94] derived optimal order quantities in the newsvendor problem under max-min loss when only mean and variance of the demand distribution are known. [LS07] used a set characterized by relative entropy distance around a nominal distribution in dynamic pricing problem. The papers [Mor59], [YW06] and [PR08] derived ordering policy using min-max regret and under different conditions on demand distribution in inventory models.

Some of the recent work on non-parametric methods that includes both learning and optimization are [LRS07], [HR09], [HLRO07] and [BZ09]. All of the above works are mainly concentrated on deriving policies that guarantee certain asymptotic performance.

In Statistics, kernel-based learning is well-known in non-parametric density estimation and non-parametric regression (see [Nad64] and [Wat64]). These non-parametric learning methods work with minimal assumptions and they typically ignore problem specific information in contrast to objective operational learning. Another key difference is non-parametric learning is primarily an estimation problem, while our task involves both decision making as well as learning. An area that is perhaps the most closely related to objective operational learning is multiple arm bandits and online learning that captures the trade off between exploration and exploitation. A long line of papers has been published on classical n-arm bandit problem and its variants starting from the seminal work of [Rob52] in stochastic setting and [ACBFS98] in adversarial case. An interesting and relevant subset of work is on online convex optimization and continuum-arm bandits. See the papers [Kle04, AK08] and [AHR08] for more details.

Unlike the work mentioned above, our estimate of the profit function incorporates useful, though possibly incorrect, assumptions about the problem when the data size is small. The advantage of making assumptions is that it reduces the variance of the objective function estimates (at the cost of introducing bias when they happen to be incorrect) which improves the small sample performance. Any structural assumption we make is smoothly dropped unlike some other approaches ([KSST08, MRT08]) where the objective function is restricted to belonging to a known (e.g. linear) subspace.

Further details on the differences between operational learning and other non-parametric approaches, as well as computational studies which show a substantial advantage of objective operational learning over non-parametric regression for small sample sizes, are provided in the body of the chapter.

6.2 Problem Description

Consider an optimization problem under uncertainty that is repeated indefinitely. At each repeat, a decision, y, is chosen from the set $\mathcal{Y} \subseteq \mathbb{R}$. Let $\phi(y)$ be the mean utility function associated with y. Our problem is to find an optimal decision y^* belonging to the optimal subset Y^*, i.e.,

$$y^* \in \mathcal{Y}^* \equiv \arg\max\{\phi(y) : \, y \in \mathcal{Y}\} \tag{6.1}$$

We consider the situations where the function $\phi(y)$ remains the same in each repeat, but is unknown to the decision maker. The optimal decision y^* can not be found by solving a single optimization problem but needs to be learned over time.

More formally, suppose that a decision (possibly randomized) Y, results in a random response X and generates an associated random utility V. In the newsvendor problem, for example, an order quantity $Y = y$ results in a realized demand X, generated from some distribution that might depend on $Y = y$, with realized profit V. Naturally,

$$\phi(y) = \mathrm{E}[V|Y = y], \qquad y \in \mathcal{Y}. \tag{6.2}$$

If Y_k denotes the decision at repeat k, with X_k and V_k the associated response and realized utility, we would like to design a method for generating Y_k (possibly randomly) given knowledge of the history of decisions, responses, and utilities $\mathcal{F}_k = \{(Y_j, \, X_j, \, V_j), \; j = 1, \ldots, \, k{-}1\}$, such that $\{Y_k, \; k = 1, \, 2, \ldots\}$ converges to an optimal decision y^*. We would like to guarantee convergence under mild assumptions about our knowledge of $\phi(y)$ and for the method to perform well for small data samples.

Note that if $Y_k = y$, $k = 1, 2, \ldots$, is the same for each repeat, then by the strong law of large numbers (SLLN), we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} V_k = \phi(y),\ a.s. \tag{6.3}$$

6.3 Non-Parametric Regression

In non-parametric statistics, kernel smoothing, introduced in [Ros56] for density estimation, plays a crucial role. Asymptotic properties of kernel smoothing are studied in [Par62] for the univariate case and in [Cac66] for the multivariate case. [Nad64] and [Wat64] extended this idea to non-parametric function regression. Therefore when ϕ is well behaved one may adapt the idea of **Nadaraya-Watson function regression** to construct a smoothed objective function as follows:

$$\tilde{\phi}(y) = \frac{\sum_{k=1}^{n} V_k \kappa_n(y, Y_k)}{\sum_{n=1}^{k} \kappa_n(y, Y_k)}. \tag{6.4}$$

Here the kernel κ_n satisfies the properties:

$$\begin{aligned} &\lim_{n\to\infty} \kappa_n(y, Y_k) = 1,\ y = Y_k, \\ &\lim_{n\to\infty} \kappa_n(y, Y_k) = 0,\ y \neq Y_k, \end{aligned} \tag{6.5}$$

and $\kappa_n(y, Y_k)$ is decreasing in $|y - Y_k|$.

When $\mathcal{Y}$ is discrete one may trivially choose $\kappa_n(y, Y_k) = 1,\ y = Y_k;\ \kappa_n(y, Y_k) = 0,\ y \neq Y_k$ for all values of k and n. When $\mathcal{Y}$ is continuous, two commonly used kernel functions are the uniform kernel

$$\left(\kappa_n(y,\, Y_k) = \mathbf{1}_{\{|y-Y_k|<\beta_n\}}\right),$$

and the Gaussian kernel

$$\left(\kappa_n(y,\, Y_k) = \exp\left(-\frac{1}{2}\frac{(y-Y_k)^2}{\beta_n^2}\right)\right),$$

where β_n ($\lim_{n\to\infty} \beta_n = \infty$) is the bandwidth parameter.

If the decisions $\{Y_k,\ k = 1, 2,\}$ are independently and uniformly sampled from $\mathcal{Y}$, then under some mild conditions on the kernel function (see the papers [Nod76] and [FY79]) one can prove:

$$\lim_{n\to\infty} \tilde{\phi}(y) = \phi(y),\ \text{(a.s., as well as, uniformly on } \mathcal{Y}). \tag{6.6}$$

If estimation of $\phi(y)$ is the only focus, this approach is adequate. However, in a decision making context, we want to choose the decisions $\{Y_k,\ k = 1, 2, \ldots\}$ so that the sequence of decisions converge to an optimal solution in an efficient way. While it is possible to devise algorithms which are completely data driven and blind to any information or knowledge about the decision problem and converge asymptotically to the optimal solution, small sample performance can be poor if domain knowledge is ignored.

6.4 Objective Operational Learning

We begin with a description of the notion of *Objective Operational Learning* and discuss its differences with the classical notion of nonparametric regression. The aim of objective operational learning is to improved finite time performance by incorporating problem-specific domain information in conjunction with data to construct the approximation $\tilde{\phi}_n(y)$ of $\phi(y)$. We illustrate these ideas in the context of the newsvendor problem.

6.4.1 Constructing the Structural model: Newsvendor problem with Observable Demand

Assume that the per-unit purchase cost of items being sold is c, the selling price is s, there is no salvage value, and demand (including lost sales) can be observed. The demand distribution may depend on the order quantity, though the precise nature of this relationship is not known to the decision maker. All repeats are statistically identical. Our data after n repeats is $\mathcal{F}_n = \{(Y_k,\, X_k,\, V_k),\ k = 1, \cdots, n\}$ where $\mathbf{Y_n} = (Y_1,\, Y_2, \ldots, Y_n)$ is the record of past decisions (order quantities), $\mathbf{X_n} = (X_1,\, X_2, \ldots, X_n)$ are the resulting observations (demands), and $\mathbf{V_n} = (V_1, \cdots, V_n)$ are the realized utilities (profits) for each repeat. Observe that $V_k = \psi(Y_k,\, X_k)$ where

$$\psi(y, x) = s \min\{y, x\} - cy,\ y \in \mathcal{Y} \equiv [y_{min}, y_{max}].$$

Our goal at each repeat is to construct an approximation of the objective function $\tilde{\phi}_n(y)$ and to use this approximation to generate a new order quantity Y_{n+1}. We would like the sequence of orders to converge to the optimal order quantity y^* as $n \to \infty$ and for the performance of the orders to be "good" even when the number of repeats is small.

An important component of the approximation that we adopt is the notion of *retrospective utility*, which as a function of y is defined as

$$\hat{\psi}_k(y,\, X_k) = \psi(y,\, X_k) = s \min\{y,\, X_k\} - cy,\ y \in \mathcal{Y}. \tag{6.7}$$

The function $\hat{\psi}_k(y,\, X_k)$ is an estimate of the profit we would have obtained in period k if y instead of Y_k happened to be the order quantity and the observed demand X_k remains unchanged even if we make this switch (hence the term *retrospective*). More generally, the function $\hat{\psi}_k(y,\, X_k)$ is a sample of the random utility associated with an order quantity y. We now consider three different cases:

(a) Standard Newsvendor Model

First, consider the special case where demands $\{X_1, X_2, \ldots\}$ are *i.i.d.* and independent of the order quantity, inventory level etc. Then $\hat{\psi}_k$ is unbiased:

$$\mathrm{E}[\hat{\psi}_k(y,\, X_k) \,|\, Y_k] = \phi(y),\; \forall\, y \in \mathcal{Y},$$

and strongly consistent:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \hat{\psi}_k(y,\, X_k) = \phi(y),\; a.s.,\; y \in \mathcal{Y},$$

which suggests the following approximation of the profit function:

$$\tilde{\phi}_n(y) = \frac{1}{n} \sum_{k=1}^{n} \hat{\psi}_k(y,\, X_k). \tag{6.8}$$

This approximation can be used to generate the decision Y_{n+1} for the next repeat. One method is to choose $Y_{n+1} \in \arg\max\{\tilde{\phi}_n(y) \,:\, y \in \mathcal{Y}\}$. In the case of *i.i.d.* and order-independent demand, it can be shown that $\tilde{\phi}_n(y) \to \phi(y)$ and $Y_n \to y^* \in \mathcal{Y}^*$ as $n \to \infty$ which is in a large part due to $\hat{\psi}_k(y,\, X_k)$ being an unbiased sample of the utility associated with order quantity y. We now consider two scenarios where $\hat{\psi}_k(y,\, X_k)$ may not be an unbiased sample of $\phi(y)$.

(b) Censored Demand Newsvendor Model

Suppose the demands are *i.i.d* and independent of the inventory level. However instead of observing full demand data only past sales are observable. The variable X_k now denotes the number of items sold in period k. Two events can occur:

- $X_k < Y_k$: sales equals demand if demand is smaller than the order quantity Y_k;
- $X_k = Y_k$: sales equals order quantity Y_k if demand exceeds Y_k.

In the censored demand case, it is not possible to say that retrospective utility $\hat{\psi}_k(y,\, X_k)$ is an unbiased sample of $\phi(y)$ for all y. However, conditioning on whether or not $y > Y_k$ we can say the following:

$$\begin{aligned} &\mathbb{E}[\hat{\psi}_k(y,\, X_k)|Y_k] &&= \phi(y) &&\text{if } y \le Y_k \\ &\mathbb{E}[\hat{\psi}_k(y,\, X_k)|Y_k] &&\ne \phi(y) &&\text{if } y > Y_k, \end{aligned}$$

i.e., the estimate $\hat{\psi}_k(y,\, X_k)$ is a biased sample of $\phi(y)$ if $y > Y_k$.

(c) Newsvendor Model with Demand Dependent on Inventory

Now suppose the demand depends on the inventory level in each repeat and X_k is demand in period k if the ordered quantity is Y_k, then the retrospective utility $\hat{\psi}_k(y,\, X_k)$, in general, will be biased in that

$$\phi(y) \neq E[\hat{\psi}_k(y,\, X_k)\,|\,Y_k], \text{ when } y \neq Y_k.$$

If we assume a functional form for the dependencies and correct the estimate $\hat{\psi}_k(y,\, X_k)$, we may have poor performance in long run if our choice of functional form turns out to be wrong.

Although the retrospective utility may be biased, the information that $\hat{\psi}_k(y,\, X_k)$ carries about the utility $\phi(y)$ can still be valuable if y and Y_k are "close", more so when there are few data points. This naturally motivates the approximation

$$\tilde{\phi}_n(y) = \frac{\sum_{k=1}^{n} \hat{\psi}_k(y,\, X_k)\kappa_n(y,\, Y_k)}{\sum_{k=1}^{n} \kappa_n(y,\, Y_k)}, \tag{6.9}$$

where $\kappa_n(y,\, Y)$ is a weighing function or a kernel. We call the approximation function $\tilde{\phi}_n(y)$, operational statistics of objective function or *Objective Operational Statistics*.

One obvious difference between Nadaraya-Watson regression (6.4) and the objective operational statistics (6.9) is that the retrospective utility $\hat{\psi}_k(y,\, X_k)$ is not used in the Nadaraya-Watson approximation, so problem specific information about the profit function captured in the retrospective utility is ignored. In addition, in contrast to equal weighting (6.8), it acknowledges the problem of bias by using the kernel to gradually suppress the influence of functions $\hat{\psi}_k(y,\, X_k)$ when Y_k is far from y.

It is also worth noting that the validity of $\hat{\psi}_k(y, X_k)$ can go beyond just the one point Y_k as we have seen in the classical newsvendor problem and in censored demand case. In such cases, the definition of the kernel used in objective operational statistics may differ from standard kernel definition used in Nadaraya-Watson regression. Note that if $\kappa_n(y, Y_k) = 1$ for all k and n, we obtain the approximation (6.8). In case of censored demand when $\hat{\psi}_k(y, X_k)$ is valid for $y \leq Y_k$, we can choose the following kernel function:

$$\begin{aligned} \kappa_n(y, Y_k) &= 1 && \text{if } y \leq Y_k \\ \kappa_n(y, Y_k) &= \kappa'_n(y, Y_k) && \text{if } y > Y_k, \end{aligned}$$

where $\kappa'_n(y, Y_k)$ satisfies the Nadaraya-Watson regression's kernel definition (6.5).

Example 15. *Consider a situation where demand is dependent on inventory level and is deterministically equal to $4 + 0.2y$. If we order Y_k in period k then the profit function is $s\min\{4 + 0.1Y_k, Y_k\} - cY_k$. Suppose $s = 4$ and $c = 2$ and we sample at points $Y_1 = 2$ and $Y_2 = 10$ and observe the corresponding profits as 4 and 0 respectively. Figure 6.1 plots the Nadaraya-Watson regression function and the objective operational statistics approximation*

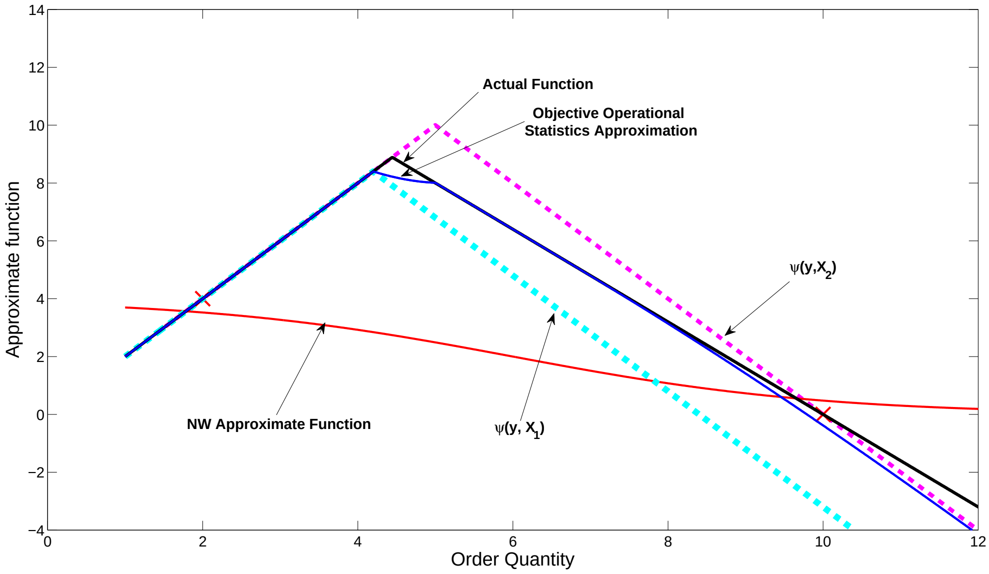

Figure 6.1: Comparison of Nadaraya-Watson regression approximation and objective operational statistics after two data points

after ordering $Y_1 = 2$ *and* $Y_2 = 10$ *and correspondingly observing demands* $X_1 = 4.2$ *and* $X_2 = 5$*. The objective operational statistics approximation is much closer to the actual function as compared to the Nadaraya-Watson regression approximation.*

6.4.2 Summary: Objective Operational Learning

A key element of the objective operational statistics approximation $\tilde{\phi}(y)$ is that it incorporates domain information through a function $\hat{\psi}_k(y, X_k)$. More generally, the information function $\hat{\psi}_k$ can be a function of all the past history, i.e., $\mathbf{X}_k = \{X_1, \ldots, X_k\}$ and $\mathbf{Y_k} = \{Y_1, \ldots, Y_k\}$. The objective operational statistics approximation in general is:

$$\tilde{\phi}_n(y) = \frac{\sum_{k=1}^{n} \hat{\psi}_k(y, \mathbf{X}_k, \mathbf{Y_k})\kappa_n(y, \mathbf{Y_k})}{\sum_{k=1}^{n} \kappa_n(y, \mathbf{Y_k})}, \tag{6.10}$$

We also need a randomized method for generating orders $\{Y_k\}$ that uses the objective operational statistics approximation while balancing exploration and exploitation trade-off. A good randomized method needs to exploit the information by sampling near points of maximum of current function estimate $\tilde{\phi}_k$, while sufficiently exploring other points to guarantee convergence of the approximation function. Our objective approximation learning algorithm which balances exploration and exploitation works as follows.

Objective Operational Learning Algorithm

0. Choose a point Y_0 arbitrarily inside $\mathcal{Y}$. Observe X_0 and construct initial objective operational estimation $\tilde{\phi}_0(y) = \psi_k(y, X_0, Y_0)$. Initialize $e(0) = 1, n = 1$ and $y_0^e = Y_0$. In addition, choose a sequence of parameters $\{\epsilon_n\}$ that control the amount of exploration.

1. Generate $U(n)$, a $\{0, 1\}$ random variable with probability $U(n) = 1$ equal to ϵ_n.

2. If $U(n) = 0$, choose $Y_n \in \mathcal{Y}_{n-1} = \arg\max\{\tilde{\phi}_{n-1}(y)\}$. In addition, $\tilde{\phi}_n = \tilde{\phi}_{n-1}$.

3. If $U(n) = 1$, increment $e(n) = e(n-1) + 1$.

4. Choose $Y_n = y^e_{e(n)}$ such that $\{y_1^e, y_2^e, \ldots, y^e_{e(n)}\}$ forms a low discrepancy sequence inside $\mathcal{Y}$. Observe X_n and define $X^e_{e(n)} = X_n$.

5. Update $\tilde{\phi}_n$ as follows:

$$\tilde{\phi}_n(y) = \frac{\sum_{k=1}^{e(n)} \hat{\psi}_k(y,\, \mathbf{X}_k^e, \mathbf{y}_k^e)\kappa_n(y,\, \mathbf{y}_k^e)}{\sum_{k=1}^{n} \kappa_n(y,\, \mathbf{y}_k^e)}, \tag{6.11}$$

where $\mathbf{y}_k^e = \{y_0^e, \ldots, y_k^e\}$ and $\mathbf{X}_k^e = \{X_0^e, \ldots, X_k^e\}$.

6. Repeat steps 2–5.

The objective operational learning algorithm balances exploration and exploitation using an $\epsilon-$greedy policy. At step n with probability $1 - \epsilon_n$ current information is exploited and Y_n is chosen corresponding to a maximum of the approximation $\tilde{\phi}_n$. With probability ϵ_n an exploration step occurs. The variable $e(n)$ denotes the number of exploration steps by time n. Note that while calculating the approximation $\tilde{\phi}_n$, only the data from exploration steps are used.

An important property of $\hat{\psi}_k$ is that it is unbiased when $y = Y_k$, namely

$$E_{X_k^e}\left[\hat{\psi}_k\Big(y,\, (\mathbf{Y}^e_{k-1},\, Y_k^e),\, (\mathbf{X}^e_{k-1},\, X_k^e)\Big) \,\Big|\, (\mathbf{Y}^e_{k-1},\, Y_k^e),\;\; \mathbf{X}^e_{k-1}\right]\Bigg|_{Y_k^e = y} = \phi(y).$$

Clearly, the art in choosing $\hat{\psi}_k$ is such that the bias is small when y is close to Y_k.

We next apply the objective operational learning algorithm to newsvendor problem with demand dependent on inventory and derive conditions on various parameters (ϵ_n, κ_n etc.) under which the functional approximation converges to the true function.

6.5 Application of Objective Operational Learning to Inventory Dependent Demand Problem

For the inventory problem described in Section 6.4.1, let $\mathcal{Y}$ denote the set of allowable order quantities and $\mathcal{D}$ the support of the random demand. We assume for each repeat that demand D, conditional on y, has cumulative distribution $F_y(\cdot)$. The expected profit associated with y is

$$\phi(y) = E[\psi(y,\, D) \,|\, y] = \int_{x \in D} \psi(y,\, x) dF_y(x).$$

The goal is to find an optimal order quantity

$$y^* = \arg\max_{y \in \mathcal{Y}} \phi(y).$$

We are interested in the situation where the distribution function $F_y(\cdot)$ (and hence $\phi(y)$) is unknown to the decision maker. Instead, the decision maker's ordering decision at repeat $n+1$ can only depend on the history of ordering decisions and the associated realizations of demand and profits $\mathcal{F}_n = \{(Y_k,\, D_k,\, V_k) \,|\, k = 1, \ldots, n-1\}$. We make the following assumptions:

Assumptions

(i) The set $\mathcal{Y}$ is bounded. Without loss of generality $\mathcal{Y} \subseteq [0, 1]$.

(ii) Profit function $\psi(y,\, D)$ is bounded on $\mathcal{Y} \times \mathcal{D}$. Without loss of generality, we assume $|\psi(y,\, D)| \leq 1$.

(iii) There is a constant $C < \infty$ such that

$$\mathbb{E}\, |\phi_{y_k}(y) - \phi(y)| \leq C\, |y - y_k|$$

where

$$\phi_{y_k}(y) = \int_{x \in D} \psi(y,\, x) dF_{y_k}(x).$$

In other words $\psi(y, D_k)$, where D_k is observed demand when $Y_k = y_k$, is a good sample of $\phi(y)$ in the sense that its expected value $\phi_{y_k}(y)$ is close to $\phi(y)$, if y is close to y_k. Note that the assumption (iii) is different than Lipschitz continuity assumption on $\phi(y)$ and the function $\phi(y)$ may even be discontinuous in y.

To implement the operational learning algorithm specified in Section 6.4.2 for inventory control problem we need to further specify the following:

- **Low discrepancy sequence.** The low discrepancy sequence $\{y_1^e, y_2^e, \ldots, y_k^e\}$ in set $\mathcal{Y}$ should satisfy the following properties:
  - For any closed interval $\mathcal{A} \subseteq [0,1]$ of length δ, there exists a constant λ^U such that

$$\sum_{i=1}^{k} \mathbf{1}_{\{y_i^e \in \mathcal{A}\}} \leq \lambda^U \max(k\delta, 1) \tag{6.12}$$

    for all $k \geq 1$.
  - For any closed interval $\mathcal{A} \subseteq [0,1]$ of length $\delta > 2/k$, there exist constants λ_L, C_L such that

$$\sum_{i=1}^{k} \mathbf{1}_{\{y_i^e \in \mathcal{A}\}} \geq \lambda_L \; k\delta \tag{6.13}$$

    for all $k \geq C_L$.

  An example of a sequence that satisfy (6.12) and (6.13) is a sequence of dyadic rationals in interval $[0,1]$. A dyadic rational is of the form $\frac{a}{2^b}$ where a and b are natural numbers. If $k = 2^b$ for some b then the k dyadic rationals are $\{0, 1/k, 2/k, \ldots, (k-1)/k\}$ and are equal-spaced. A dyadic rational sequence can be constructed as follows: Let $b_m b_{m-1} \ldots b_3 b_2 b_1$ be the binary representation of $k-1$. Then the k^{th} dyadic rational y_k^e has a binary representation $0.b_1 b_2 b_3 \ldots b_{m-1} b_m$. It is easy to see that the sequence of dyadic rationals satisfy (6.12) and (6.13) with $\lambda^U = 1, \lambda_L = 1/2, C_L = 1$.

- **Function $\hat{\psi}_k(y, \mathbf{y}_k^e, \mathbf{X}_k^e)$.** Note that observed data $\mathbf{X}_k^e$ is demand data $\{D_1^e, \ldots, D_e^k\}$ corresponding to points $\{y_1^e, \ldots, y_e^k\}$. We make a simplifying assumption that the function $\hat{\psi}_k(y, \mathbf{y}_k^e, \mathbf{X}_k^e) = \psi(y, D_k^e)$. For convergence results we do not use any specific form of ψ (hence our results can be extended to other OR problems). Specific form of $\psi(y, D)$, for example, $\psi(y, D) = s\min(y, D) - cy$, where s and c are selling price and cost of purchase respectively, can be used while implementing the algorithm. The update equation corresponding to (6.11) is

$$\tilde{\phi}_n(y) = \frac{\sum_{k=1}^{e(n)} \psi(y,\, D_k^e)\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}. \tag{6.14}$$

- **Kernel Function.** We use a uniform kernel with bandwidth β_n, i.e.,

$$\kappa_n(y, y_k^e) = \mathbf{1}_{\{|y - y_k^e| < \beta_n\}} \tag{6.15}$$

  Even though we prove convergence results for uniform kernel, extending convergence results to other kernels such as Gaussian kernel is straightforward.

- **Sequence $\{\epsilon_n\}$.** First,note that $e_n = \sum_{i=1}^{n} U_n$, where U_n is independent Bernoulli random variable with probability of success ϵ_n. Let $b_n = \sum_{i=1}^{n} \epsilon_i$. We choose the sequence $\{\epsilon_n\}$ in such a way that $b_n \geq \log n$ eventually. Following fact is immediate using Chernoff bound for sum of independent Bernoulli random variables:

  Proposition 16. *If $b_n \in \omega(\log n)$, then for any $\delta > 0$*

$$(1-\delta)b_n \leq e(n) \leq (1+\delta)b_n \quad \textit{eventually, almost surely.}$$

We now prove the point-wise convergence of the function $\tilde{\phi}(y)$.

Theorem 17. *Suppose assumptions (i)-(iii) are satisfied and the sequence $\{y_1^e, y_2^e, \ldots\}$ satisfies the condition (6.13). If the bandwidth parameter β_n in kernel function (6.15) is chosen to be $\beta_n = \alpha b_n^{-\frac{1}{3}}, \alpha > 0$, where $b_n \in \omega(\log n)$ and $\tilde{\phi_n}(y)$ is updated according to (6.14), then the following holds for every $y \in Y$:*

$$\limsup_{n\to\infty} \mathbb{E}\left[\frac{(\tilde{\phi}_n(y) - \phi(y))^2}{b_n^{-\frac{2}{3}}}\right] \leq M < \infty,$$

where M is a constant. The rate of convergence of $\tilde{\phi}_n(y)$ in expectation is therefore $b_n^{-\frac{1}{3}}$.

Proof.

$$\begin{aligned}
(\tilde{\phi}_n(y) - \phi(y))^2 &= \left(\frac{\sum_{k=1}^{e(n)} \psi(y,\, D_k^e)\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} - \phi(y)\right)^2 \\
&= \left(\frac{\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}\right. \\
&\quad \left. + \frac{\sum_{k=1}^{e(n)} (\phi_{y_k^e}(y)) - \phi(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}\right)^2 \qquad (6.16) \\
&\leq 2\left(\frac{\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}\right)^2 \qquad (6.17) \\
&\quad + 2\left(\frac{\sum_{k=1}^{e(n)} (\phi_{y_k^e}(y)) - \phi(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}\right)^2
\end{aligned}$$

If $\kappa_n(y,\, y_k^e) = 0$ when $|y - y_k^e| > \beta_n$ and therefore, using assumption (iii)

$$\begin{aligned}
\left(\frac{\sum_{k=1}^{e(n)} (\phi_{y_k^e}(y)) - \phi(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 &\le \left(\frac{\sum_{k=1}^{e(n)} C|y_k^e - y|\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 \\
&\le \left(\frac{\sum_{k=1}^{e(n)} C\beta_n\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 \\
&= C^2\beta_n^2. \qquad (6.18)
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{E} \left(\frac{\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 \\
&\quad = \mathbb{E} \left[\mathbb{E} \left[\left(\frac{\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 \middle| e(n) \right] \right] \\
&\quad = \mathbb{E} \left[\frac{\mathbb{E} \left[\left(\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e) \right)^2 \middle| e(n) \right]}{\left(\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e) \right)^2} \right]. \qquad (6.19)
\end{aligned}$$

Using the fact that $(\psi(y,\, D_i^e) - \phi_{y_i^e}(y))$ and $(\psi(y,\, D_j^e) - \phi_{y_j^e}(y))$ are independent mean 0 random variables when $i \ne j$ and $\mathbb{E}(\psi(y,\, D_i^e) - \phi_{y_i^e}(y))^2 \le 4$ using assumption (ii) we obtain:

$$\begin{aligned}
\mathbb{E} \left(\frac{\sum_{k=1}^{e(n)} (\psi(y,\, D_k^e) - \phi_{y_k^e}(y))\kappa_n(y,\, y_k^e)}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right)^2 &\le \mathbb{E} \left[\frac{\sum_{k=1}^{e(n)} 4\kappa_n^2(y,\, y_k^e)}{\left(\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e) \right)^2} \right] \qquad (6.20) \\
&= 4\mathbb{E} \left[\frac{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)}{\left(\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e) \right)^2} \right] \qquad (6.21) \\
&= 4\mathbb{E} \left[\frac{1}{\sum_{k=1}^{e(n)} \kappa_n(y,\, y_k^e)} \right] \qquad (6.22) \\
&\le 4\mathbb{E} \left[\frac{1}{\lambda_L e(n)\beta_n} \right] \le \frac{4}{\lambda_L b_n(1-\delta)\beta_n}, \qquad (6.23)
\end{aligned}$$

where (6.21) follows from the fact that $\kappa_n^2(\cdot\ ,\ \cdot) = \kappa_n(\cdot\ ,\ \cdot)$ as the kernel value is either 0 or 1. Equation (6.23) follows from (6.13) and proposition 16. Now from (6.17), (6.18) and (6.23)

$$\mathbb{E}\left[(\tilde{\phi}_n(y) - \phi(y))^2\right] \leq \frac{8}{\lambda_L(1-\delta)b_n\beta_n} + 2C^2\beta_n^2. \tag{6.24}$$

For the choice of $\beta_n = \alpha b_n^{-1/3}$,

$$\mathbb{E}\left[(\tilde{\phi}_n(y) - \phi(y))^2\right] \leq M\ b_n^{-2/3}, \tag{6.25}$$

where $M = \frac{8}{\alpha\lambda_L(1-\delta)} + 2C^2\alpha^2$. □

Chapter 7

Conclusion

In this thesis, we investigated and studied some important issues in modeling uncertainty with learning.

First, we considered the problem of worst-case robust intensity control of the arrival and departure processes of a single-state queuing system, where model ambiguity is represented using the notion of relative entropy. The queuing problem belongs to the class of robust dynamic optimization problems. A novel feature of our model is that we consider different levels of uncertainty for the arrival and departure processes. We prove that the optimal robust control for our model is of threshold type. Our work on queuing control work can be extended in several directions. One possibility is to consider state-dependent capacity limits. Another extension is to consider multistage networks. Finally it would be interesting to consider a decentralized version of our problem where arrivals and departures are controlled by separate entities. Our approach should also extend to other intensity control problems such as the ones that arise in optimal dynamic pricing of goods.

When limited amount of past data is available for learning, a popular approach in robust optimization is to construct uncertainty sets for unknown parameters from past data using confidence sets. We showed that using such an approach may results in a policy which has worse performance than a non-robust policy. We introduced a new approach called generalized operational statistics and applied it to a mean-variance portfolio optimization problem. The operational statistics policy for portfolio optimization guarantees a better solution than a classical sample mean based policy. In our work we have assumed that only the mean returns vector is uncertain while the covariance matrix is known with certainty. An important extension of our work would be a portfolio optimization problem, such as minimum variance portfolio problem, where the covariance matrix is assumed to be uncertain. In addition, the policy class we considered is restrictive in the sense that if the uncertainty set is a singleton, then we do not obtain the optimal policy corresponding to the singleton. Because of this, the performance of worst case robust policy for mean variance portfolio optimization outperforms the operational statistics policy for small uncertainty sets.

The generalized operational statistics approach we introduced in this thesis should be applicable to many operations research problems. One difficulty in applying operational statistics approach to broader range of problems is the numerical tractability of the solutions. With advances in the field of convex optimization, specially with development of interior point algorithms for second order conic programs and semidefinite programs, it is now possible to solve many large scale robust problems efficiently. In portfolio optimization we exploited the structure of the operational statistics problem and proved that the operational statistics portfolio optimization problem can be converted into a semidefinite program and therefore, can be solved efficiently. In general, it is quite possible that a operational statistics is not convex even if the nominal problem is convex. However, it is worth exploring the class of problems where the operational statistics formulation remains convex.

Finally, we introduced a new relatively model free learning approach called objective operational learning. As opposed to other non-parametric approaches, objective operational learning incorporates structural information about the objective and the problem, and therefore aims for a good small sample performance while guaranteeing convergence to the true solution under stationarity of underlying stochastic process. In most real world processes, the stationarity assumption is not likely to be valid, and hence it is very important to look at algorithms which have good small sample performance. We showed how to effectively incorporate structural information in objective operational learning algorithm by demonstrating several examples of newsvendor problems. We applied the objective operational learning algorithm to an inventory control problem with demand distribution dependent on inventory level and proved various structural properties.

The objective operational statistics approach would be very useful in high dimensional setting such as multi-product pricing, where model free approaches that do not utilize any structural information would take a long time to learn. In high dimensions, the small sample performance of classical non-parametric strategy is expected to be poor. In high dimensional setting, one can incorporate multiple levels of structural information into operational statistics algorithm. For example, in multi-product problems, demand distributions of two products can be stochastically ordered. Another example is a case where the demand of a particular product is always less than the sum of demands of two other products. We would also have the information about the profit function which may be a complex non-linear function of product prices and their demands. A multi-dimensional objective operational learning algorithm should be designed to incorporate as much information as possible into the algorithm.

Bibliography

[ACBF02] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.

[ACBFS98] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: The adversarial multiarmed bandit problem. *In Proceedings of the 36th Annual FOCS.*, 1998.

[AHR08] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. *In 21st Annual Conference on Learning Theory.*, 2008.

[AK08] B. Awerbuch and R. Kleinberg. Online linear optimization and adaptive routing. *Journal of Computer and System Sciences*, 74(1):97–114, 2008.

[Azo85] K. S. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31:1150–1160, 1985.

[Bar74] C. B. Barry. Portfolio anlysis under uncertain means, variances and covariances. *Journal of Finance*, 29:515–522, 1974.

[BB01] J. Baxter and P.L. Bartlett. Infinite-horizon gradient-based policy search. *Journal of Artificial Intelligence Research*, 15:319–350, 2001.

[BBK79] V. S. Bawa, S. Brown, and R. W. Klein. Estimation risk and optimal portfolio choice: A survey. *Proceedings of the Americal Statistical Association*, pages 53–58, 1979.

[Ber85] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, NY, 2 edition, 1985.

[Ber95] D.P. Bertsekas. *Dynamic Programming and optimal Control Vol II.* Athena Scientific, Belmont, Massachusetts, 1995.

[BM02] V. S. Borkar and S. P. Meyn. Risk sensitive optimal control for Markov decision processes with monotone cost. *Mathematics of Operations Research*, 27(1):192–209, 2002.

[Bre81] P. Bremaud. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York, 1981.

[Bro93] M. Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45:21–58, 1993.

[BS04] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[BTN98] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.

[BTN99] A. Ben-Tal and A. Nemirovski. Robust solutions to uncertain programs. *Operations Research Letters*, 25:1–13, 1999.

[BTN00] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[BTN01] A. Ben-Tal and A. Nemirovskii. *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[BTar] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, to appear.

[BZ09] O. Besbes and A. Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

[Cac66] T Cacoullos. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18:179–189, 1966.

[CCdO03] R. Cavazos-Cadena and R. Montes de Oca. The value iteration algorithm in risk-sensitive average Markov decision problems with finite state space. *Mathematics of Operations Research*, 28:752–776, 2003.

[CdMK06] W. L. Cooper, T. Homen de Mello, and A. J. Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54:968–987, 2006.

[CE02] Z. Chen and L. G. Epstein. Ambiguity, risk and asset returns in continuous time. *Econometrica*, 70:1403–1443, 2002.

[Cho93] V. K. Chopra. Improving optimization. *Journal of Investing*, 8:51–59, 1993.

[CM99] S. P. Coraluppi and S. I. Marcus. Risk sensitive and minimax control of discrete-time, finite-state Markov decision processes. *Automatica*, 35:301–309, February 1999.

[CSS08] L. Y. Chu, J. G. Shanthikumar, and Z. J. M. Shen. Solving operational statistics via a bayesian analysis. *Operations Research Letters*, 36:110–116, 2008.

[CY90] H. Chen and D. Yao. Optimal intensity control of a queuing sysytem with state-dependent capacity limit. *IEEE Transactions on Automatic Control*, 35(4), 1990.

[CZ93] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.

[DGNU09] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, May 2009.

[DGU09] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy. *Review of Financial Studies*, 22:1915–1953, 2009.

[Dur03] R. Durrett. *Probability: Theory and Examples*. Duxbury, 2003.

[EL97] L. El Ghaoui and H. Lebret. Robust solutions to least-square problems with uncertain transition matricies. *Siam J. Matrix Anl. Appl.*, 18:1035–1064, 1997.

[Ell61] D. Ellsberg. Risk, ambiguity and savage axioms. *Quarterly Jornal of Economics*, 75:643–669, 1961.

[EOO03] L. El Ghaoui, M. Oks, and F. Oustry. Worst case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[EW94] L. G. Epstein and T. Wang. Intertemporal asset pricing under knightian uncertainty. *Econometrica*, 62:283–322, 1994.

[FS88] P. A. Frost and J. E. Savarino. For better performance constrain portfolio weights. *Journal of Portfolio Management*, 15:29–34, 1988.

[FY79] Schuster E. F. and S. Yakowitz. Contributions to the theory of nonparametric regression, with application to system identification. *Annals of Statistics*, 7:139–149, 1979.

[Gal92] G Gallego. A minimax distribution-free procedure for the Q R inventory model. *Operations Research Letters*, 11:55–60, 1992.

[GI03] D. Goldfarb and G. Iyengar. Robust portfolio slection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[GM94] G. Gallego and I. Moon. The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society*, 44(8):825–834, 1994.

[H58] H. *Studies in the Mathematical theory of Inventory and Production*. Stanford University Press, 1958.

[HHM96] D. Hernandez-Hernandez and S. I. Marcus. Risk sensitive control of Markov processes in countable state space. *Systems and Control Letters*, 29:147–155, November 1996.

[HLRO07] W. T. Huh, R. Levi, P. Rusmevichientong, and J. Orlin. Adaptive data-driven inventory control policies based on kaplan-meir estimator. 2007.

[HR09] W. T. Huh and P. Rusmevichientong. A non-parametric asymptotic analysis on inventory planning with censored demand. *Mathematics of Operations Research*, 34:103–123, 2009.

[HSTW06] L.P. Hansen, T.J. Sargent, G.A. Turmuhambetova, and N. Williams. Robust control and model misspecification. *Journal of Economic Theory*, 128(1):45–90, May 2006.

[Jac79] J. Jacod. Calcul stochastique et problèmes de martingales. In *Lecture Notes in Math. 714*, New York, 1979. Springer-Verlag, Berlin Heidelberg.

[Jef46] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461, September 1946.

[JK81] J. Jobson and B. Korkie. Putting Markowitz Theory to Work. *Journal of Portfolio Management*, 7:70–74, 1981.

[JM03] R. Jagannathan and T. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1684, 2003.

[Jor86] P. Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21:279–292, 1986.

[JS61] W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the 4th Berkeley Symposium on Probability and Statistics 1*. University of California Press, 1961.

[Kle04] R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 17:697–704, 2004.

[KLM96] L. P. Kaelbling, M.L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[Kni21] F. H. Knight. *Risk, Uncertainty, and Profit.* Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA, 1921.

[KSST08] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari. Efficient bandit algorithms for online multiclass prediction. *Proceedings of the 25th International Conference on Machine Learning*, 2008.

[Li88] L. Li. A stochastic theory of the firm. *Mathematics of Operations Research*, 13:447–466, 1988.

[LRS07] R. Levi, R. O. Roundy, and D. B. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32:821–839, 2007.

[LS91] L. Lovász and A. Schrijver. Cones of Matrices and Setfunctions, and 0-1 Optimization. *SIAM Journal of Optimization*, 1(2):166–190, 1991.

[LS05] L. Liyanage and J. G. Shanthikumar. A practical inventory policy using operational statistics. *Operations Research Letters*, 33:341–348, 2005.

[LS07] A.E.B. Lim and J.G. Shanthikumar. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research*, 55(2):198–214, 2007.

[LSW09] A.E.B. Lim, J.G. Shanthikumar, and T. Watewai. Robust multi-product pricing and a new class of risk-sensitive control problems. *Working Paper*, 2009.

[LW04a] O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix: Problems in mean-variance. *Journal of Portfolio Management*, 30:110–119, 2004.

[LW04b] O. Ledoit and M. Wolf. Well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411, 2004.

[Mar52] H. M. Markowitz. Portfolio Selection. *Journal of Finance*, 7:77–91, 1952.

[Mar59] H. M. Markowitz. *Portfolio Selection.* Wiley, New York, 1959.

[MG94] I. Moon and G. Gallego. Distribution free procedures for some inventory models. *Journal of Operations Research Society*, 45(6):651–658, 1994.

[Mic89] R. O. Michaud. The markowitz optimization enigma: Is optimized optimal. *Financial Analysts Journal*, 45:31–42, 1989.

[Mor59] W. T. Morris. Inventorying for unknown demand. *Journal of Industrial Engineering*, 10(4):299–302, 1959.

[MP00] A. C. MacKinlay and L. Pastor. Asset pricing models: Implications for expected returns and portfolio. *The Review of Financial Studies*, 13:883–916, 2000.

[MRT08] A. J. Mersereau, P. Rusmevichientong, and J. N. Tsitsiklis. A structured multi-armed bandit problem and the greedy policy. *47th IEEE Conference on Decision and Control*, pages 4945–4950, 2008.

[MS01] J. Moody and M. Saffell. Learning to trade via direcft reinforcement. *IEEE Transactions on Neural Networks*, 12(4), 2001.

[Nad64] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9:141–142, 1964.

[Nl05] A. Nilim and l. El ghaoui. Robust solutions to Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[NN94] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[Nod76] K. Noda. Estimation of a regression function by the parzen kernel-type density estimators. *Journal of the Institute of Mathematical Statistics*, 28(2):221–234, 1976.

[Par62] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1067, 1962.

[PJD00] I. R. Petersen, M. R. James, and P. Dupuis. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control*, 45:398–412, 2000.

[PMR96] P. Dai Pra, L. Meneghini, and W. J. Runggaldier. Connections between stochastic control and dynamic games. *Mathematics of Control, Signals, and Systems*, 9:303–326, 1996.

[PR08] G. Perakis and G. Roels. Regret in the newsvendor problem with partial information. *Operations Research*, 56(1):188–203, 2008.

[Rob52] H. Robbins. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58:527–535, 1952.

[Ros56] M. Rosenblatt. Remarks on some nonparametric estimates of density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

[SB98] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[Sca58] H. E. Scarf. A min-max solution to an inventory problem. In S. Karlin K. J. Arrow and H. E. Scarf, editors, *Studies in Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, Stanford, CA, 1958.

[Ser81] R. Serfozo. Optimal control of random walks, birth and death processes, and queues. *Advances in Applied Probability*, 13:61–83, 1981.

[Sha53] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39:1095–1100, 1953.

[Ste55] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *3rd Berkeley Symposium on Probability and Statistics 1*, pages 197–206. University of California Press, 1955.

[Sti85] S. Stidham. Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, 30:705–713, 1985.

[Sut88] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44, 1988.

[Sut90] R. S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In Morgan Kaufmann, editor, *Proceedings of the Seventh International Conference on Machine Learning*, 1990.

[TK04] R. Tütüncü and M. Koenig. Robust asset allocation. *Annals of Operations Research*, 132:157–187, 2004.

[UW03] R. Uppal and T. Wang. Model misspecification and under diversification. *Journal of Finance*, 58:2465–2486, 2003.

[VB93] L. Vandenberghe and S. Boyd. Primal–dual potential reduction method for problems involving matrix inequalities. Technical report, Stanford, CA 94305, USA, January 1993.

[Wat64] G. S. Watson. Smooth regression analysis. *Shankya: The Indian Journal of Statistics, Series A*, 26:359–372, 1964.

[Wat89] C. J. C. H. Watkins. *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK, 1989.

[Wri97] S. Wright. *Primal-dual interior point methods*. SIAM, Philadelphia, PA, 1997.

[YW06] J. B. Chen Yue and M. C. Wang. Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136, 2006.