SEMIPARAMETRIC AND NONPARAMETRIC METHODS IN DATA MINING AND STATISTICAL LEARNING WITH APPLICATIONS IN PUBLIC HEALTH SURVEILLANCE AND PERSONALIZED MEDICINE

Yingqi Zhao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2012

Approved by:

Michael R. Kosorok
Donglin Zeng
Amy H. Herring
David B. Richardson
Wei Wang

Abstract

YINGQI ZHAO: SEMIPARAMETRIC AND NONPARAMETRIC METHODS IN DATA MINING AND STATISTICAL LEARNING WITH APPLICATIONS IN PUBLIC HEALTH SURVEILLANCE AND PERSONALIZED MEDICINE
(Under the direction of Michael R. Kosorok)

The field of statistical learning has been growing rapidly over the past few decades, with a diverse range of applications. In this dissertation, we develop methodology mainly using semiparametric and nonparametric statistical learning techniques for the areas of public health surveillance and personalized medicine.

Surveillance, providing early warning for impending emergencies, is a key function of public health. In Chapter 2, we propose a semiparametric spatiotemporal method to model spatiotemporal lattice data via a local linear fitting combined with day-of-week effects, in which both spatial and temporal information are taken into account. Detection of abnormal events are carried out using an ARMA time series technique for residuals combined with a resampling approach to determine the threshold for significance. We conduct simulations to assess the performance of the proposed method. Also, the method is illustrated using the data on daily asthma admissions collected through North Carolina emergency departments that occurred between 2006 and 2007.

There is increasing interest in personalized medicine: the idea of tailoring treatment for each individual to optimize patient outcome. In Chapter 3, we focus on the single-decision setup. We show that estimating such an optimal treatment rule is equivalent to a classification problem where each subject is weighted proportional to his or her clinical outcome, although the true class labels, to which treatment group the patients belong as the optimal, are unknown in the training set. We then propose a new approach based on the support vector machine framework from computer science. We show the resulting estimator of the treatment rule is consistent, and further derive fairly accurate convergence rates for this estimator. The performance of the proposed approach is demonstrated via simulation studies and an analysis of chronic depression data.

It is not uncommon that the best clinical strategies may require adaptation over time. We thus in Chapter 4 generalize the outcome weighted learning method to the multi-decision setup, aiming at finding the dynamic treatment regimes, customized sequential decision rules for individual patients which can adapt over time to the evolving illness, to maximize the long term health outcome. Inspired by the intrinsic idea in dynamic programming, we conduct outcome weighted learning for each stage backwards through time. We further introduce an iterative procedure which can improve the performance of the algorithm. The methods are evaluated by simulation studies and an analysis on a smoking cessation data set.

Acknowledgments

From the depth of my heart I express my deepest gratitude to my advisor, Dr. Michael R. Kosorok, for his guidance in developing statistical methods. His constant encouragements and generous financial support during my PhD studies have helped me walk through the entire process. His passion, commitment and disciplined focus into research and teaching, have been consistently inspiring me to strive for excellence and become an independent scholar. I deem it as my privilege to work under his direction.

I would like to give sincere thanks to my committee members: Dr. Donglin Zeng, Dr. Amy H. Herring, Dr. David B. Richardson and Dr. Wei Wang. My special thanks goes to Dr. Donglin Zeng. His knowledge and insights in the biostatistical field have largely enriched my understanding of the discipline. His resourceful suggestions helped me successfully complete my dissertation. I appreciate Dr. Amy H. Herring for her perceptive comments about my research and continuous encouragements throughout my graduate years at UNC. I am thankful to Dr. David B. Richardson, who helped me develop and improve my collaborative and interdisciplinary research skills. I owe my thanks to Dr. Wei Wang for providing helpful suggestions in improving my dissertation.

I am deeply grateful to Dr. Jianwen Cai for kindly helping me identify the initial research assistant opportunity. Also, working with her for the last two years at UNC in Translational and Clinical Sciences Institute was an invaluable experience for me. I want to thank Dr. Anna Waller, Amy Ising, Dr. Eric Laber and Dr. A. John Rush for their constructive suggestions on my work.

I record my appreciation to faculties and staff from the Department of Biotatistics for providing a supportive and comfortable working environment, and valuable suggestions to achieve my goal. I do thank all the members in the Kosorok research group for their interesting and stimulating discussions. My heartfelt thanks to all my friends, who have stood beside me, through the hard times and the good, during my graduate studies.

Last but not the least, I am forever indebted to my family for their endless love, support and encouragement throughout my life.

Table of Contents

List of Tables x

List of Figures xi

List of Abbreviations xii

1 Introduction 1

1.1 Overview of Statistical Learning 1

1.2 Overview of Public Health Surveillance 3

1.3 Overview of Personalized Medicine 4

1.3.1 Individualized Treatment Rules for Single-Decision Setup 5

1.3.2 Dynamic Treatment Regimes for Multi-Decision Setup 6

1.4 Outline of Thesis 7

2 Detecting Disease Outbreaks Using Local Spatiotemporal Methods 9

2.1 Introduction 9

2.2 Methodology 13

2.2.1 Model 13

2.2.2 Estimation 14

2.2.3 Detecting Outbreaks via Residual Analysis 19

2.3 Data Application and Simulation Study 22

2.3.1 Data Application: ED Asthma Admissions 22

2.3.2 Simulation Studies 24

2.4 Discussion 26

3 Estimating Individualized Treatment Rules Using Outcome Weighted Learning 32

3.1 Introduction 33

3.2 Methodology 34

3.2.1 Individualized Treatment Rule (ITR) 34

3.2.2 OWL for Estimating Optimal ITR 35

3.2.3 Linear Decision Rule for Optimal ITR 37

3.2.4 Nonlinear Decision Rule for Optimal ITR 39

3.3 Theoretical Results 40

3.3.1 Notation 41

3.3.2 Fisher Consistency 41

3.3.3 Excess Risk for $\mathcal{R}(f)$ and $\mathcal{R}_\phi(f)$ 42

3.3.4 Consistency and Risk Bounds 43

3.3.5 Improved Rate with Data Completely Separated 46

3.4 Simulation Study 47

3.5 Data Analysis 51

3.6 Discussion 53

4 Optimal Dynamic Treatment Regimes using Outcome Weighted Learning 58

4.1 Introduction 58

4.2 General Methodology 63

4.2.1 Dynamic Treatment Regimes (DTR) 63

4.2.2 Outcome Weighted Learning for the Multi-Decision Setup . . . 66

4.3 Theoretical Results 72

4.3.1 Fisher Consistency 72

4.3.2 Consistency 73

4.3.3 Risk Bound and Convergence Rate 74

4.3.4 Improved Rate with Data Completely Separated 77

4.4 Simulation Studies 78

4.5 Data Analysis: Smoking Cessation Study 81

4.6 Discussion 83

5 Future Work 94

5.1 Identifying Potential Causes in Public Health Surveillance 94

5.2 Extensions in Optimal DTRs Discovery 95

5.3 Concluding Remarks 97

Appendix 1: Chapter 3 Proofs 98

Appendix 2: Chapter 4 Proofs 111

Bibliography 115

List of Tables

2.1 Proportion of Detected Outbreaks: Constant Pattern 27

2.2 Proportion of Detected Outbreak: Peak Pattern 28

2.3 Number of False Outbreaks Detected 28

3.1 Mean Depression Scores 55

4.1 Mean Values of the Estimated DTR, Scenario 1 85

4.2 Mean Values of the Estimated DTR, Scenario 2 85

4.3 Mean Values of the Estimated DTR, Scenario 3 85

4.4 Proportions of Higher or Same Values 86

4.5 Baseline and Intermediate Variables 86

4.6 Mean Cross Validated Values using Different Methods 86

List of Figures

2.1 Day-of-Week Effects 29

2.2 Q-Q Plot of Residuals 29

2.3 Population Distance for North Carolina 30

2.4 Baseline Weekly Asthma Rate 31

2.5 Counties in NC Detected for Potential Outbreaks 31

3.1 Contour Plots of Value Function 55

3.2 MSE for Value Functions 56

3.3 MSE for Misclassification Rates 57

4.1 Kernel Density Approximations, Scenario 1 87

4.2 Q-Q Plot of Estimated Values for Scenario 1 88

4.3 Kernel Density Approximations, Scenario 2 89

4.4 Q-Q Plot of Estimated Values for Scenario 2 90

4.5 Kernel Density Approximations, Scenario 3 91

4.6 Q-Q Plot of Estimated Values for Scenario 3 92

4.7 Q-Q Plot of Cross Validated Values 93

List of Abbreviations

BOWL	Backwards outcome weighted learning
CBASP	Cognitive Behavioral-Analysis System of Psychotherapy
CUSUM	Cumulative sums
DTR	Dynamic treatment regimes
ED	Emergency departments
HRSD	24-item Hamilton Rating Scale for Depression
IOWL	Iterative outcome weighted learning
ITR	Individualized treatment rules
MDD	Major depressive disorder
NC	North Carolina
NC DETECT	North Carolina Disease Event Tracking and Epidemiologic Collection Tool
OLS	Ordinary least squares
OWL	Outcome weighted learning
PLS	Penalized least squares
Q-Q	Quantile-Quantile
RKHS	Reproducing kernel hilbert space
SMART	Sequential multiple assignment randomized trial
SUTVA	Stable unit treatment value assumption
SVM	Support vector machine

Chapter 1

Introduction

In this dissertation, we investigate two problems: (1) the problem of developing a reliable on-line surveillance system which can quickly identify anomalies and provide early warning for impending emergencies (Chapter 2); and (2) the problem of discovering the optimal individualized treatment strategies for different patients based on their own characteristics, which can improve patients outcomes on average if the discovered strategies are implemented in the future (Chapter 3 for single-decision setup and Chapter 4 for multi-decision setup). Both research problems are studied using statistical learning combined with semi- and non-parametric modeling techniques.

1.1 Overview of Statistical Learning

Over the last few decades, the field of machine learning has been growing at an unprecedented rate. It is usually categorized into supervised learning and unsupervised learning. In supervised learning, we have a training set of data where outcomes and a set of features are observed on subjects. The task is to build a prediction model, with which we can predict the values of the outcomes for new subjects. In the unsupervised learning problem, we observe only features without outcomes being measured. In this situation, we concentrate on the structures of the observed feature data, for example, how they are organized or clustered. Given a dataset with observations recorded, a machine learning algorithm builds a model based on the data, and generalizes to future data using the estimated model.

All the learning procedures are implemented by the computer without human intervention in machine learning. Statistical learning, on the other hand, formulates the learning methods within the probabilistic framework. Statistics enables a rigorous analysis of machine learning methods and provides guarantees on the expected results. The prediction task in supervised learning is called regression when the outcomes are continuous, and classification when the outcomes are discrete. Statistical decision theory provides a framework for developing models that can fulfill the prediction task, which requires a loss function to penalize the prediction error. A natural choice for continuous outcomes is the squared error loss functions, while zero-one loss is applicable for categorical outcomes. Measure of success can be evaluated via the expectation of the loss function over the joint distribution of the data, which leads to a clear understanding in adequacy and effectiveness of different methods. Common prediction methods include parametric learning methods (linear model fit by least squares possibly along with shrinkage methods), semi-parametric learning methods (neural networks), non-parametric learning methods (k-nearest-neighbor prediction rule, kernel smoothing methods, tree based methods, support vector machines and boosting methods). Unsupervised learning focuses on directly inferring the properties of the probability density of the feature data. Most commonly used techniques include association rules, cluster analysis, principal components and the variants, independent component analysis. Interested readers can refer to Hastie et al. (2009) for more details.

An important subfield of the non-traditional setting, with stochastic sequential decision processes involved, is referred as reinforcement learning. In this context, there is an “agent”, which is a learner or decision-maker. “Environment” refers to the thing the agent interacts with, comprising everything outside the agent. While the agent continually interacts with the environment, it produces a sequence of “actions”, to which the environment can respond and provide feedback in turn. As a consequence, the agent can achieve a goal, usually to maximize the total amount of reward it receives over the long run, by learning from its own experience. More details can be found in Sutton and Barto (1998).

Statistical learning has been widely applied to various areas related to biostatistical research, for example, investigating the influences of prognostic factors on the clinical outcomes using the regression approach (Pages 49-51, Hastie et al. (2009)), classifying biological samples using gene expression data (Tibshirani et al., 2002), treating behavioral disorders with reinforcement learning (Pineau et al., 2007). In this dissertation, we develop methodology using semiparametric and nonparametric statistical learning techniques, which can be applied in the fields of public health surveillance and personalized medicine.

1.2 Overview of Public Health Surveillance

The role of public health surveillance is to collect, analyze, and interpret data related to planning and evaluation of public health practice, and to disseminate the information to administrators (Thacker and Berkelman, 1988). Specifically, off-line surveillance monitors the process retrospectively for some fixed and predetermined time period. On-line systems monitor processes continually observed in real time, and try to detect the aberrations as quickly as possible after they occur. Conventionally, public health surveillance relies on off-line analysis (Ogata, 1988; Kulldorff, 1997; Schoenberg, 1999, 2003; Gangnon and Clayton, 2004; Kulldorff et al., 2005). Though traditional off-line disease surveillance approaches for the early detection of outbreaks can offer a closer and more reliable supervision, the utility of it is limited by delays in obtaining and analyzing the data. The recent interest in surveillance requires the analysis be done in near real time. This has motivated a large literature concerning anomaly detection in the on-line situation. A variety of temporal anomaly detection methods have been developed for the purpose of surveillance (Hutwagner et al., 1997; Lewis et al., 2002; Tsui et al., 2003; Reis and Mandl, 2003). However, pure temporal surveillance methods are not sufficient when we can collect space and time data, since they lack power to detect outbreaks starting locally or they can result in severe problems of multiple testing if carried out on several small areas simultaneously. It is more appropriate to do space-time detection, incorporating available spatial information (Kulldorff, 2001; Kleinman et al., 2004; Diggle et al., 2005; Karr et al., 2009). The complexity in public health surveillance necessitates the development of new methodologies to handle statistical issues involved. In the first part of this dissertation, we introduce a statistical model and computational methods for disease surveillance and illustrate the approach by showing how the method helps with alarm detection in a timely manner.

1.3 Overview of Personalized Medicine

In many different diseases, patients can show significant heterogeneity in response to treatments. Among multiple active treatments which are available, a treatment that works for a majority of individuals may not work for a subset of patients with certain characteristics. The emerging field of personalized medicine, enabling a more personalized approach to health care, has been offering possibilities for improving the health of individuals. It is worth noting that newly developed drugs may be abandoned because no significant improvements have been detected across a population, whereas it is highly possible that subgroups of patients could be benefit from them. Thus the goal of personalized medicine is to achieve the optimal clinical outcome by steering patients to the right drug at the right dose at the right time (Hamburg and Collins, 2010). In this case, the decisions and practices are tailored for the individual patients, based on all the information from them, including but not limited to demographics information, clinical measures, medical histories and genetic information. A complete process for effective personalized medicine discovery typically includes five key steps (Ren et al., 2012): obtain patients individual data that can reflect personal characteristics; identify potential biomarkers that may indicate the stratification of patients into subgroups with common features (Eisen et al., 1998; Dhanasekaran et al., 2001; Yeung and Ruzzo, 2001; Tibshirani et al., 2002); develop and select candidate therapeutic regimens; measure the relationship between clinical outcomes and prognostic variables, such as biomarkers, and treatment choices; and verify the relationship in a prospective randomized clinical trial. Specifically, challenges arising from discovering effective treatment regimes by estimating the relationship between outcomes and individual predictors have motivated us to develop new methodologies to tackle the wide range of potential statistical problems. In the second and third part of this dissertation, we propose methods for finding the optimal treatment assignment rule based on individual characteristics within a non-parametric statistical learning framework.

1.3.1 Individualized Treatment Rules for Single-Decision Setup

Very often existing methods for assigning treatments make the assumption that patients are homogeneous. It is important to recognize that a better understanding of patient heterogeneity will lead to greater health outcomes. There has been a considerable amount of literature focusing on this matter. For example, molecularly targeted cancer drugs are only effective for patients with tumors expressing targets (Grünwald and Hidalgo, 2003; Buzdar, 2009), and a high variability exists in responses among patients with different levels of psychiatric symptoms (Piper et al., 1995; Crits-Christoph et al., 1999). Ishigooka et al. (2000) found that some patients with schizophrenia may experience remarkable improvement, others may not or even have worsening symptoms after taking olanzapine. Fukuoka et al. (2011) show that epidermal growth factor receptor (EGFR) mutations are the strongest predictive biomarker for administrating first-line therapy in patients with advanced non-small cell lung cancer, that progression free survival and tumor response were significantly improved for gefitinib versus chemotherapy. Thus significant improvements in public health could potentially result from judiciously treating individuals based on his or her prognostic or genomic data rather than using a “one size fits all”approach.

Treatments and clinical trials tailored for patients have enjoyed recent popularity in clinical practice and medical research, and, in some cases, have provided high quality recommendations accounting for individual heterogeneity (Sargent et al., 2005; Insel, 2009). These proposals have focused on smaller, specific and well-defined subgroups, sought to provide guidance in clinical decision making based on individual differences, and have attempted to achieve better risk minimization and benefit maximization.

1.3.2 Dynamic Treatment Regimes for Multi-Decision Setup

It is not uncommon that the best clinical strategies may require adaptation over time. Recognizing that there exist time varying characteristics among patients, and moreover, that the nature of diseases is as evolving and diversified as the people, clinicians have found that treatments which work now may not work later. This is especially common in the case of chronic diseases. Even if a “once and for all dosing” is easy to implement, it may not be the best strategy for the patients. To name a few, treatment for major depressive disorder is usually driven by additional factors emerging over time, such as side-effect severity, treatment adherence and so on (Murphy et al., 2007); typically, the regimen for cancer patients involve multiple lines of treatment to improve survival, for example, non small cell lung cancer (Socinski and Stinchcombe, 2007);

clinicians routinely update therapy according to the risk of toxicity and antibiotics resistance in treating cystic fibrosis (Flume et al., 2007). Such problems have motivated a vast literature on personalized treatment strategies. Ideally, treatment decisions should adapt with time dependent outcomes, such as patients response to previous treatments and side effects. Moreover, instead of focusing on a short-term benefit of a treatment, the goal should be an improvement of the long-term gain by considering the treatment delayed effects to the patients.

Dynamic treatment regimes (DTR), also called adaptive treatment strategies (Murphy, 2005a), are sequential decision rules for individual patients which can adapt over time to an evolving illness. At each decision point, the covariate and treatment histories of a patient are taken as input for the decision rule, which outputs an individualized treatment recommendation subsequently. Therefore, not only heterogeneities among individuals are taken into consideration, but also those across time within an individuals are incorporated into this framework. In other words, various aspects of treatment strategies, including treatment types, dosage levels, timing of delivering and etc, can evolve with time according to subject-specific needs. On the other hand, it has drawn the attention of researchers that treatments resulting in the best immediate effect may not necessarily lead to the most favorable long term outcomes. Consequently, with the flexibility of managing the long-term clinical outcomes, dynamic treatment regimes have become increasingly popular in clinical practice. In general, the goal is to identify the optimal dynamic treatment regime, defined as the rule that will maximize the mean response at the end of the time period.

1.4 Outline of Thesis

In the present chapter, we have reviewed some of the existing literature regarding statistical learning. We have also introduced the general concepts and problems of interests in the fields of public health surveillance and personalized medicine. In Chapter 2, we develop a real-time surveillance system via local linear fitting and residual analysis. In Chapter 3, we propose a novel outcome weighted learning framework to estimate the optimal individualized treatment rules within the single-decision setup, where the clinician only makes one time decision for the patient. The derived regimes tailored to different patients can maximize the expected outcomes if they are implemented on future patients. It is likely that the clinician has several decision times to determine the treatment. Thus we need to identify customized sequential decision rules for individual patients which can adapt over time to the evolving illness. The proposed outcome weighted learning methodology is generalized to optimal dynamic treatment regimes discovery within the multi-decision setup in Chapter 4. We discuss possible extensions and future work in Chapter 5.

Chapter 2

Detecting Disease Outbreaks Using Local Spatiotemporal Methods

In this chapter, a real-time surveillance method is developed with emphasis on rapid and accurate detection of emerging outbreaks. See Zhao et al. (2011b). We develop a model with relatively weak assumptions regarding the latent processes generating the observed data, ensuring a robust prediction of the spatiotemporal incidence surface. Estimation occurs via a local linear fitting combined with day-of-week effects, where spatial smoothing is handled by a novel distance metric that adjusts for population density. Detection of emerging outbreaks is carried out via residual analysis. Both daily residuals and AR model-based de-trended residuals are used for detecting abnormalities in the data given that either a large daily residual or an increasing temporal trend in the residuals signals a potential outbreak, with the threshold for statistical significance determined using a resampling approach.

2.1 Introduction

The primary purpose of disease surveillance is to detect unusual spatial or temporal patterns of disease; this may lead to further investigation to determine the causes of an unusual pattern of disease. To provide early warning and enable rapid public health intervention, it is important to develop a reliable disease surveillance system that can quickly identify anomalies. Such a system might combine available information sources such as emergency department (ED) visit data, physician office data, disease reports from clinical laboratories, and data on over-the-counter drug sales.

The North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) was created by the NC Division of Public Health (NC DPH) in 2004 in collaboration with the UNC Department of Emergency Medicine to provide statewide early recognition of outbreaks and monitoring of public health using various data sources. Currently, NC DETECT's sources of data include emergency departments (ED), the Carolinas Poison Center, and the Pre-hospital Medical Information System (PreMIS), as well as pilot data from the NCSU College of Veterinary Medicine Laboratories and select urgent care centers. All hospitals in NC with 24-hour-acute care emergency departments must report electronic data in near-real time. This dissertation focuses on the use of spatiotemporal methods for early identification and situational awareness of unexpected variation in the incidence of asthma reported to NC DETECT from emergency departments, considering ED admissions that occurred between January 1, 2006 and December 31, 2007. A variety of methods have been developed for the purpose of surveillance. A standard surveillance tool used by CDC, as well as NC DETECT, is cumulative sums (CUSUM), a quality control method (Hutwagner et al., 1997). CUSUM accumulates deviations between observed and expected values; if the expectation is modeled poorly, signals detected by CUSUM may not reflect the true underlying changes. Time series approaches are also used to detect disease outbreaks (Reis and Mandl, 2003; Craigmile et al., 2007), although these methods do not account for spatial correlation. Openshaw et al. (1987) developed a graphical method, the geographical analysis machine (GAM), counting the observed cases in multiple overlapping circles with increasing radii and identifying those significantly higher than the expected value. GAM, however, is criticized by Kulldorff and Nagarwalla (1995) for its limitation in handling multiple testing. Kulldorff (1997) proposed a spatial scan statistic, widely used for cluster investigations, which searches over a given set of regions and applies the likelihood ratio test for the null hypothesis that the probability of events happening in the window is the same as that outside the window. However, it is difficult to derive the exact distribution of this likelihood ratio statistic. Bernoulli and Poisson data are analyzed using the spatial scan statistic in Kulldorff (1997), and it has also been generalized to ordinal data (Jung et al., 2007). Cucala et al. (2009) note that the spatial scan statistic can be computationally infeasible, proposing instead graph-based spatial scan tests linking those events closer than a given distance, allowing completely data-based clusters rather than only those of predetermined shape. This has been generalized to the detection of disease clusters in the space-time domain (Kulldorff et al., 1998). Application of prospective disease surveillance may be done using space-time scan statistics to detect new clusters resulting from an emerging disease outbreak (Kulldorff, 2001; Neill et al., 2005).

Representing the cases as a point pattern, it is natural to extend point process methodology to the surveillance problem. There is a vast literature on spatial point processes, motivated by growing numbers of data sets being collected in fields such as epidemiology, environmental studies, geography, seismology and forestry (Ripley, 1977; Cressie, 1993; Møller and Waagepetersen, 2004). However, methodology for the analysis of spatiotemporal point processes is less well developed.

One common approach for the analysis of spatiotemporal point processes is to consider the conditional intensity, which can be interpreted as the hazard of the occurrence of an event at location s and time t, given the history of the process over the interval $[0, t]$, which uniquely determines the probability structure of the point process when it exists (see Daley and Vere-Jones (2003)). Estimation of this intensity function, however, is more important in scenarios for which the mean behavior of the process is of primary interest, and estimates for the intensity function are usually obtained parametrically, see, e.g. Schoenberg (2004). However, because the assumptions made for model based inference may not be valid, Diggle et al. (1985) proposed a nonparametric estimator utilizing kernel smoothing, selecting bandwidths via data-driven procedures.

Diggle et al. (2005) formulated the problem of online spatiotemporal disease surveillance by obtaining predictions in the context of a non-stationary log-Gaussian Cox process and calculating the exceedance probabilities of the intensities over a pre-specified threshold value. Unfortunately, unreliable estimates may result from violations of the Cox model assumption.

We develop a real-time surveillance method, utilizing a local linear estimation method incorporating day-of week effects to construct the predicted spatiotemporal incidence rate surfaces and implementing residual analysis to identify anomalies. Our method makes fewer assumptions about the latent processes, resulting in greater robustness in prediction than existing methods. We do not impose Poisson process or log Gaussian Cox process (Diggle et al., 2005) assumptions where intensity functions are straightforward for modeling; instead, we only assume event counts at each location follow a Poisson distribution with a marginal Poisson rate. We first form a model to estimate marginal Poisson rates, which is complex enough to capture the spatiotemporal structure of the data; the proposed prediction methodology accounting for spatial and temporal variation in the underlying incidence rates enables us to detect true clusters more precisely. We can then carry out flexible testing for abnormalities based on residual analysis in a computationally fast manner. The surveillance system we develop has usefulness beyond cluster detection because users can gain information on the regular pattern of incidence, which other methods (e.g. scan statistics) cannot provide.

In section 2.2 we propose a spatiotemporal model for the marginal Poisson rate that is estimated using local linear regression methods. We then discuss detection of outbreaks via residual analysis. In Section 2.3 we present a data application using asthma ED admissions to illustrate how the proposed surveillance system works and provide results of a carefully designed simulation study comparing our method to the CUSUM and space-time scan statistic (Kulldorff, 2001). Finally, we summarize the results and give a brief discussion in Section 2.4.

2.2 Methodology

2.2.1 Model

Let the number of observed cases at location s on day t be $N(s,t)$, and let $n(s)$ be the population of location s, which we assume is constant during the study period. We want to provide a reliable prediction of the normal pattern of spatial and temporal incidence of cases, i.e., we need to estimate the incidence rates for the underlying spatiotemporal point process from the collected data.

When studying events such as hospital admissions or emergency department visits, there are some commonly-observed systematic patterns to event occurrence. For example, the number of ED admisssions at a hospital tends to be greater on weekends than on week days. Therefore, natural day-of-week variation must be accounted for in the model. We assume that the observed incidence rate at a given location and on a given day is the product of a base incidence rate at this location and a value dependent on the day of week. The model for the Poisson rate $\lambda(s,t)$ is:

$$\lambda(s,t) = \mu(s,w)e^{\sum_{l=1}^{7}\alpha_l(s)I(\text{Day}=l)} + \epsilon(s,t), \tag{2.2.1}$$

subject to

$$\sum_{l=1}^{7} \alpha_l(s) = 0$$

where $\mu(s, w)$ is the baseline Poisson rate of location s in the week w in which day t, the l^{th} day of the week, occurs. (i.e. for Sunday, January 8, $t = 8, w = 2, l = 1$.) We assume the day-of-week effect, denoted by $\alpha_l(s), l = 1, \cdots, 7$ is the same across the whole period but varies across the region, allowing each location to have a specific weekly pattern. Then we estimate the expected rate at location s on day t and use this to detect the outbreaks.

2.2.2 Estimation

The model (2.2.1) has two types of parameters: local parameters for the weekly baseline effect $\mu(s, w)$ and global parameters for the day-of-week effect $\alpha_l(s)$. Local parameters are only determined by a local neighborhood, while global parameters depend on all locations. We discuss the estimation procedures in the following sections.

Local Linear Estimates for Baseline Weekly Effect

The weekly baseline rate is the geometric mean of each day's rate in the given week. We focus on the smoothing method for incidence rates by day first and then discuss the estimation of day-of-week effects.

Nonparametric smoothing methods have been widely used in many statistical areas, although they have seen relatively little use in the point process field. Diggle et al. (1985) discusses kernel smoothing methods for one-dimensional point processes. The weaknesses of this type of kernel smoother have been well discussed, including its poor boundary performance, large bias and low efficiency (see e.g. Fan and Gijbels (1996)). The boundary effects can be even worse in two or higher dimensions. We therefore smooth the three-dimensioned spatiotemporal point processes via local linear fitting, which is gaining increasing popularity for its desirable properties over kernel smoothing.

We implement local linear fitting by approximating the unknown Poisson rate function λ by a linear function in time in the neighborhood of any point (s, t), using weighted least squares, with weights provided by a spatiotemporal kernel. For location s, we estimate $\lambda(s, w_k)_l$, $l = 1, \cdots, 7$, which is the rate in location s for the l^{th} day of the k^{th} week, denoted by w_k, in a year, based on that day's data (e.g, all Mondays).

For every l^{th} day of the week, at each (s, t_l), we solve

$$\min \sum_{i,k} K_{i,k}(s, t_l)(\lambda(s_i, t_{w_k l}) - \beta_0(s, t_l) - \beta_1(s, t_l)(t_{w_k l} - t_l))^2, \tag{2.2.2}$$

for $\hat{\beta}_0(s, t_l), \hat{\beta}_1(s, t_l)$. Here i indexes the location, $i = 1, \cdots, S$, and k indexes the weeks, $k = 1, \cdots, M$. We have

$$K_{i,k}(s, t_l) = K_1\left(\frac{\|s - s_i\|}{c_{s,t_l} a_n}\right) K_2\left(\frac{\|\theta(t_l) - \theta(t_{w_k l})\|}{a_n}\right), \tag{2.2.3}$$

where $\lambda(s_i, t_{w_k l}) = N(s_i, t_{w_k l})/n(s_i)$, $N(s_i, t_{w_k l})$ is the recorded visit count on the l^{th} day of k^{th} week, denoted by $t_{w_k l}$ at location s_i, and $n(s_i)$ is the population of location s_i. $K(s, t_l)$ is the kernel function assigning weights to points in a region around (s, t_l) based on the distance from (s, t_l). The kernel function is factored into a spatial kernel K_1 and a temporal kernel K_2, accounting for spatial and temporal effects on the incidence rate. We use a Gaussian kernel function for both kernels. The weight decreases exponentially with the distance in the spatial domain from s in K_1 and in the temporal domain from t_l in K_2. The temporal distance is defined based on the function $\theta(t) = \exp(i2\pi t/365)$. We also define a new distance metric—population density adjusted distance—in the space kernel, and we leave this for later discussion.

The solution to the minimization problem (2.2.2) is easily shown to be $(X'WX)^{-1}X'W\Lambda$, where Λ is a vector of length ST with stacked rates, and W is an $ST \times ST$ diagonal matrix with $(((7k+l)-1)S+i)^{th}$ entry being

$$K_1\left(\frac{\|s-s_i\|}{c_{s,t_l}a_n}\right)K_2\left(\frac{\|\theta(t)-\theta(t_{w_k l})\|}{a_n}\right),$$

where T is the total number days during the study period. a_n is a sequence of bandwidths tending to zero as n goes to infinity. Note that a_n controls the size of the local temporal neighborhood. Because of different scales in space and time, we use a scaling factor c_{s,t_l} in the kernel K_1, where

$$c_{s,t_l}=\frac{\text{median}_i\|s-s_i\|}{\text{median}_j\|\theta(t_l)-\theta(t_{w_j l})\|}.$$

This simplifies the selection of two bandwidths simultaneously in a reasonable way because c_{s,t_l} reflects the relative magnitude in the spatial domain compared to the temporal domain. By multiplying a_n by c_{s,t_l}, we can have the bandwidth controlling the size of the spatial neighborhood. The bandwidths can then be selected via data-driven cross validation. Thus we obtain

$$\hat{\lambda}(s,t_l)=\hat{\beta}_0(s,t_l). \tag{2.2.4}$$

Using all the estimated daily incidence rates, we can now calculate baseline weekly effects as

$$\hat{\mu}(s,w_k)=\sqrt[7]{\prod_{l=1}^{7}\hat{\lambda}(s,t_{w_k l})}, \tag{2.2.5}$$

where $\hat{\lambda}(s,t_{w_k l})$ is obtained from (2.2.4).

Temporal Distance

For two time points $t_l, t_{w_{j\,l}}$, the length of the time interval $|t_l - t_{w_{j\,l}}|$ is a common distance measure. We propose instead to use the temporal distance $|\theta(t_l) - \theta(t_{w_{j\,l}})|$, where $\theta(t) = \exp(i2\pi t/365)$, ensuring the rates at the end of one year and beginning of the next should be close, as this complex number distance metric has attractive features in representing cyclical patterns.

Population density adjusted distance

Distance measures widely used in spatial data analysis include ordinary Euclidean distance, great circle distance, and others. We are interested in the disease cluster being adjusted for spatial variations in the population density because potential clustering can emerge by chance due solely to population clustering. There is information about case incidence in the population, and intuitively, more densely populated areas contain more information than rural areas. We propose a new distance measure for spatial smoothing that adjusts for population density in the sense that we try to stretch out the denser area and make sparser areas more compact, in order to ensure a similar degree of smoothing for all individuals in the population.

Let D_i denote the population density in the i^{th} location. Let ρ_g be the Euclidean distance measure. We propose an adjusted distance between two locations s_1, s_2 defined as

$$\rho_p(s_1, s_2) = \sum_{i=1}^{S} \int_0^1 \frac{1}{h} \tilde{K}\left(\frac{\rho_g(s_1 + u(s_2 - s_1), s_i)}{h}\right) D_i d(\rho_g(s_1 + u(s_2 - s_1), s_i)). \quad (2.2.6)$$

We use a kernel smoothing method with a Gaussian kernel $\tilde{K}$. Let $v_i(u) = \rho_g(s_1 +$

$u(s_2 - s_1), s_i)$, minimized globally at u_i. Define weights w_j as follows: if $u_i \in [0,1]$,

$$w_i = \Phi\left(\frac{v_i(0)}{h}\right) + \Phi\left(\frac{v_i(1)}{h}\right) - 2\Phi\left(\frac{v_i(u_i)}{h}\right);$$

otherwise if $u_i \notin [0,1]$,

$$w_i = \Phi\left(\frac{\max\{v_i(0), v_i(1)\}}{h}\right) - \Phi\left(\frac{\min\{v_i(0), v_i(1)\}}{h}\right).$$

Thus we have

$$\rho_p(s_1, s_2) = \sum_{i=1}^{S} w_i D_i.$$

The proposed distance between two locations $\rho_p(s_1, s_2)$ is actually a weighted sum of all the locations' population densities, while the weights depend on the real geo-distances between s_1 and s_2. Compared to the geo-distances commonly used in spatial statistics, the adjusted distance has appealing properties in terms of spatial smoothing in surveillance applications. Figure 2.3 illustrate Euclidean distance versus population density adjusted distance for the state of North Carolina.

Estimation of Day-of-Week Effect

The day-of-week effect is handled in an ANOVA-type framework. Given that we have the estimated baseline effects $\hat{\mu}(s, w_k)$ from (2.2.5), we can use these in our objective function (2.2.1). Let $\alpha(s_i) = (\alpha_1(s_i), \cdots, \alpha_7(s_i))'$, and $U_{w_k}^l = (I(t_{w_k l} = \text{Sun}), \cdots, I(t_{w_k l} = \text{Sat}))'$. Then

$$\hat{\alpha}(s_i) = \arg\min \sum_{i,k,l} (\lambda(s_i, t_{w_k l}) - \hat{\mu}(s_i, w_k) e^{\alpha(s_i)^T U_{w_k}{}^l})^2.$$

We can estimate $\hat{\alpha}(s_i)$ using the Fisher scoring algorithm as follows. Letting

$$Q(\alpha(s_i)) = \sum_{i,k,l} (\lambda(s_i, t_{w_k l}) - \hat{\mu}(s_i, w_k) e^{\alpha(s_i)^T U^l_{w_k}})^2,$$

we have

$$\begin{aligned}
\frac{\partial Q(\alpha(s_i))}{\partial \alpha(s_i)} &= -2 \sum_{i,j} U^l_{w_k} \hat{\mu}(s_i, w_k) e^{\alpha(s_i)^T U^l_{w_k}} (\lambda(s_i, t_{w_k l}) - \hat{\mu}(s_i, w_k) e^{\alpha(s_i)^T U^l_{w_k}}), \\
\frac{\partial^2 Q(\alpha(s_i))}{\partial \alpha(s_i)^2} &= -2 \sum_{i,k,l} U^l_{w_k} {U^l_{w_k}}^T \hat{\mu}(s_i, w_k) e^{\alpha(s_i)^T D_j} \lambda(s_i, t_{w_k l}) \\
&+ \; 4 \sum_{i,k,l} U^l_{w_k} {U^l_{w_k}}^T \hat{\mu}(s_i, w_k)^2 e^{2\alpha(s_i)^T U^l_{w_k}}, \quad \text{and} \\
\mathrm{E}\left(\frac{\partial^2 Q(\alpha(s_i))}{\partial \alpha(s_i)^2} \right) &= 2 \sum_{i,k,l} U^l_{w_k} {U^l_{w_k}}^T \hat{\mu}(s_i, w_k)^2 e^{2\alpha(s_i)^T U^l_{w_k}},
\end{aligned}$$

subject to

$$\sum_{l=1}^{7} \alpha_l(s) = 0.$$

Then we obtain

$$\alpha(s_i)^{n+1} = \alpha(s_i)^n + \left\{ \mathrm{E}\left(-\frac{\partial^2 Q(\alpha(s_i)^n)}{\partial \alpha(s_i)^2} \right) \right\}^{-1} \left\{ \frac{\partial Q(\alpha(s_i)^n)}{\partial \alpha(s_i)} \right\}.$$

Figure 2.1 plots the average day-of-week effects on asthma rate in North Carolina. With respect to admissions to the ED, the plot indicates a general higher rate on Sunday and then followed by Monday, although this effect varies for different counties.

2.2.3 Detecting Outbreaks via Residual Analysis

Once we observe new data, i.e, the case records from each location, we calculate the residuals for detection of abnormalities in the data. A square root transformation is applied to the count data to stabilize the variance, with the residual for the disease rate defined as

$$X_{ij} = 2\left(\sqrt{\frac{N(s_i, t_j)}{n(s_i)}} - \sqrt{\hat{\lambda}(s_i, t_j)}\right),$$

where j indexes the days in the study period, $j = 1, \cdots, T$. This raw residual is taken after removing the systematic pattern in the counts, adjusted for spatial effects.

Each location has a residual time series of high autocorrelation. Fitting an AR(2) model to the i^{th} residual series, we have

$$X_{ij} = \mu_i + \phi_{i1}X_{i,j-1} + \phi_{i2}X_{i,j-2} + \epsilon_{ij}. \tag{2.2.7}$$

Our diagnostic shows that the autocorrelation vanishes after AR(2) model fitting in some degree for our data, and the residuals from the model are approximately white noise, see Figure 2.2. Moreover, the model residual ϵ_{ij} is taken after the trend components in X_{ij} are removed. This de-trended model residual, as opposed to the raw residual, can better reflect temporal changes in the raw residual. It is possible that other time series models should be fitted when analyzing data of different diseases or from different areas, but the idea summarized in this section is generic. Given that we can estimate μ_i, ϕ_{i1} and ϕ_{i2} from (2.2.7), the model residual is

$$e_{ij} = X_{ij} - \hat{\mu_i} - \hat{\phi}_{i1}X_{i,j-1} - \hat{\phi}_{i2}X_{i,j-2}.$$

We then have two types of residuals: original/raw residuals, X_{ij}, and model based residuals, ϵ_{ij}, for location s_i on day t_j. Both residuals are useful for detection, because either a statistically significant large daily residual or a large temporal change in the residuals can indicate a possible disease outbreak at a specific time and county. We conduct inference based on the standardized residuals, which are compared on the same scale, as well as the unstandardized residuals which are less variable. In the NC DETECT data, we frequently encounter higher variability in incidence rates for locations with a larger population base. The unstandardized residuals are more sensitive in these scenarios, but the tests based on standardized residuals can have lower power to detect outbreaks.

We use a resampling approach to determine the threshold for significance. This permits detecting abnormalities without being overly sensitive. We outline the procedures below:

1. From the time series model in (2.2.7), we estimate $\hat{\mu}_i, \hat{\phi}_{i1}, \hat{\phi}_{i2}$ and $\hat{\sigma}_i^2$.

2. Critical values for e_{ij} are then obtained by

    - Generating replicates $Z_i \sim MVN(0, \hat{\Sigma}_{S\times S}), i = 1, \cdots, m$, where $\Sigma_{S\times S} = diag(\hat{\sigma}_i^2)$.
    - For each vector Z_i, taking the maximum $\gamma_i = \max_j |Z_{i,j}|$ and $\gamma_i' = \max_j |Z_{i,j}|/\sigma_i$. Let $\Gamma = (\gamma_1, \cdots, \gamma_m)$ and $\Gamma' = (\gamma_1', \cdots, \gamma_m')$. This step is mainly done to control the type I error rate. We obtain the critical value for the unstandardized model filter residuals, obtaining $\Gamma_{(0.95m)}$, and for the standardized ones, obtaining $\Gamma'_{(0.95m)}$, which are the 95% quantiles of Γ and Γ'.

3. Critical values for X_{ij}s are obtained as follows. To generate raw residual sequences, we need to retain the first and the second true X_ts as baselines. The steps are

    - Generate $Z_{i3} \sim N(0, \hat{\sigma}_i^2), i = 1, \cdots, S$.
    - Calculate $X_{i3} = \hat{\mu}_i + \hat{\phi}_{i1} X_{i2} + \hat{\phi}_{i2} X_{i1} + Z_{i3}, i = 1, \cdots, S$.
    - Among all the generated $X_{\cdot 3}$, take $\delta_1 = \max_i |X_{i3}|$ for the unstandardized and $\delta_1' = \max_i |X_{i3}|/\sigma_i$ for the standardized versions.
    - Repeat the previous steps m times, and record δ_i and δ_i' each time. Let $\Delta = (\delta_1, \cdots, \delta_m)$ and $\Delta = (\delta_1', \cdots, \delta_m')$. We obtain the critical value for the

unstandardized raw residuals, denoted $\Delta_{(0.95m)}$, and for the standardized ones, denoted $\Delta'_{(0.95m)}$.

Assuming the critical value for raw residuals is c_1, and the critical value for model residuals is c_2 (the model filtered critical value), we have

$$2\left(\sqrt{\frac{N(s_i,t_j)}{n(s_i)}} - \sqrt{\hat{\lambda}(s_i,t_j)}\right) \le c_1,$$

which leads to

$$\frac{N(s_i,t_j)}{n(s_i)} \le \times \left(c_1/2 + \sqrt{\hat{\lambda}(s_i,t_j)}\right)^2. \qquad (2.2.8)$$

and, for the model filtered critical value, we have

$$\frac{N(s_i,t_j)}{n(s_i)} \le \left(\frac{\hat{\mu_i} + \hat{\phi}_{i1}X_{i,j-1} + \hat{\phi}_{i2}X_{i,j-2} + c_2}{2} + \sqrt{\hat{\lambda}(s_i,t_j)}\right)^2. \qquad (2.2.9)$$

While using the critical value of the standardized version—denoted c_3 for the scaled raw residuals and c_4 for the scaled model filter residuals—we replace c_1 with $c_3\hat{\sigma}_i$ in (2.2.8) and c_2 with $c_4\hat{\sigma}_i$ in (2.2.9). This is very informative in the sense that users can have direct impressions about how large the disease case counts are for determining emerging outbreaks.

2.3 Data Application and Simulation Study

2.3.1 Data Application: ED Asthma Admissions

The data of interest consist of daily counts of asthma ED admissions. For every patient, data on age, sex, county of residence, date of admission to the emergency department, and diagnosis at discharge (coded according to the 9th Revision of the International Classification of Diseases Clinical Modification) are recorded. Mid-year population estimates for 2006 and 2007 were derived using the Population Division of the United States Bureau of the Census county population estimates for NC.

The surveillance system we propose requires that we develop a predictive model for the outcome of interest. Prediction is carried out using existing data for prior days of observation. In our example, we predict daily asthma incidence rates for each county in NC using the prior 2 years of NC DETECT data; as the system is currently configured, we update this predictive model every 90 days in order to incorporate more recent data. At each update, county-specific critical values are obtained and tabled for use in determining whether daily counts are excessive over the next 90 day period.

For the prediction, we need to obtain the population density adjusted distance measure for NC as described in Section 2.2.3. We want to make the local window narrow in order to obtain stable smoothed distances adjusted for the population density, choosing the bandwidth for the kernel in (2.2.6) to be 0.1 mile. Figure 2.3 shows the implied North Carolina county layout based on our proposed distance.

The baseline weekly incidences are obtained following the local spatio-temporal approaches discussed in Section 2.2.1. When doing daily incidence estimation, we let cross validation automatically choose the bandwidth a_n in (2.2.3) from 0.241, 0.361, 0.482, 0.602 and 0.723, corresponding to $\theta(14), \theta(21), \theta(28), \theta(35)$ and $\theta(42)$. In other words, we vary the temporal window from 2 weeks to 6 weeks, and cross validation chooses the best fitted one. Figure 2.4 shows the baseline weekly ED asthma admission rate of the 9^{th}, 22^{nd}, 35^{th} and 48^{th} weeks of the year for all the counties of North Carolina. These weeks are shown mainly because they announce the beginning of spring, summer, autumn and winter. Our prediction reflects the seasonal variation in asthma. As can be seen from the figure, the whole state experiences higher incidence in asthma as spring approaches, yet the rates are reduced in the summer season and increase when fall comes. This is expected, considering that allergies, particularly prevalent in the spring and fall due to high pollen levels in the air, can have a huge impact.

Following the strategies described in Section 2.3, we carry out the residual analysis to identify aberrations from regular patterns. All the counties identified to have potential outbreaks during the two-year period are shown in Figure 2.5, which can give users a direct view of the these counties' distribution.

The map suggests that patterns in the years 2006 and 2007 are very similar: outbreaks are more likely emerging at the northeastern and western boundaries of North Carolina. This may be due to regional differences in pollution levels, weather trends and population demographics. It could also be due to variations in access to health care outside the ED in these regions of the state. Lack of access to good primary care may mean poor management of the disease, more or more severe exacerbations, and lack of options for seeking care in a crisis.

2.3.2 Simulation Studies

We create simulated data sets by artificially producing outbreaks in the current data. We want to see if our method can detect emerging outbreaks rapidly and accurately, comparing it to CUSUM and scan statistics.

We chose two counties: Wake, population size 832,875, and Hyde, population size 5449, for adding counts; they represent large and small counties with no known outbreaks in 2006-2007. One simulated outbreak period starts on November 10^{th}, 2006, and the other starts on April 10^{th}, 2007. The prediction of counts for Hyde County during the simulated outbreak periods ranges from 0.03 to 0.49, and for Wake County ranges from 10.56 to 16.70. Generally, we obtain higher admission counts estimates in the spring than in the fall.

We incorporate two types of outbreaks for the simulation: one with a high constant

Poisson intensity, and the other with a peak intensity in the middle of the period, with intensity increasing beforehand and decreasing afterward. We consider both short (3 days) and long (7 days) lasting outbreaks to compare different surveillance algorithms' sensitivities. The counts in the simulated outbreak with constant intensity follow a Poisson distribution with rate λ, $P(\lambda)$. To simulate patterns with a peak, we set the starting date and have the counts initially follow a $P(\lambda_0)$ distribution. During the following days, we first increase and then decrease the intensity by λ_1 every day. Therefore, for example, if we want to generate an outbreak with a peak seven days long, the corresponding intensities for the counts generated are λ_0, $\lambda_0 + \lambda_1$, $\lambda_0 + 2\lambda_1$, $\lambda_0 + 3\lambda_1$, $\lambda_0 + 2\lambda_1$, $\lambda_0 + \lambda_1$, and λ_0. We first present the proportions of simulated outbreaks being detected, defined as any day in the simulated period having an alarm, out of 500 simulation runs.

As illustrated in Table 2.1 and Table 2.2, we see that CUSUM has extremely high sensitivity at the expense of sounding excessive false alarms. While flagging possible aberrations is of priority for a surveillance system, we are hesitant to apply the CUSUM algorithm due to its oversensitivity from so many detections. For the scan statistics, we did not conduct as many simulations due to computational complexity. Instead, we randomly selected 5 simulated data sets for each pattern-period combination, and tested them using scan statistics. Our study suggests that the scan statistic is not sensitive when detecting short time outbreaks, especially for small counties. For example, it fails to detect any aberration for three-day simulated periods, except when the intensity for Wake is as high as 70. Scan statistics identify one cluster including more than 20 counties, and the length of the outbreak is from September to December, which is too general for the public health system to take prompt corresponding action.

While our method generally has lower sensitivity than the CUSUM method, it also drastically reduces the number of false alarms as shown in Table 2.3, providing a reasonable algorithm for real-time outbreak detection. Thus we found the scan statistic to be insensitive to detecting short term outbreaks, the CUSUM method to have strong sensitivity at the expense of sounding excessive alarms (on average two thousand false alarms in a 2-year period in NC), and our method to have acceptable sensitivity and a greatly reduced false positive rate.

2.4 Discussion

The local spatiotemporal estimation method is here applied to aggregated data over counties, i.e., the number of cases in the counties that reported in local EDs. The proposed approach can be naturally extended to point process data. By modeling the space-time intensity of incident cases from accumulated historical data, we can predict the regular pattern if there are no outbreaks. Inference can further be done based on the sptiotemporal residuals obtained after removing the normal trends.

In disease surveillance, there are often many different data streams or different outcomes that we want to monitor. Kulldorff et al. (2007) propose an extension of the spatial and space-time scan statistic that can simultaneously handle multiple data sets. We want to develop techniques to monitor many potential health outcomes simultaneously, such as obesity, asthma, stroke, and unintentional injury, based on our existing work. This can be very challenging, given that we will encounter a fairly complicated situation of having correlated health outcomes in addition to the existing complex spatiotemporal dependence structure.

The new surveillance method will be further developed and validated. It is expected that the new methodology will be generally useful for conducting state-wide or nation-wide public health surveillance.

Table 2.1: Proportion of Simulated Outbreaks being Detected: Constant Pattern

Three Days Outbreak with Constant λs							
Hyde County		λ					
		1	2	3	4	5	6
Fall,2006	Proposed Methods	65%	92%	100%	100%	100%	100%
	CUSUM	95%	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	8%	33%	85%	95%	98%	100%
	CUSUM	97%	98%	100%	100%	100%	100%
Wake County		λ					
		30	40	50	60	70	80
Fall,2006	Proposed Methods	0%	11%	75%	100%	100%	100%
	CUSUM	99%	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	10%	81%	99%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%	100%
Seven Days Outbreak with Constant λs							
Hyde County		λ					
		1	2	3	4	5	6
Fall,2006	Proposed Methods	77 %	96%	99%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	14%	60%	89%	99%	100%	100%
	CUSUM	100%	100%	100%	100%	100%	100%
Wake County		λ					
		30	40	50	60	70	80
Fall,2006	Proposed Methods	0%	13%	89%	100%	100%	100%
	CUSUM	97%	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	21%	95%	100%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%	100%

The start date of outbreak for fall, 2006 is November 10^{th}, and the start date of outbreak for spring, 2007 is April 10^{th}. Each outbreak lasts for 3 days. Simulated count in Hyde (Wake) County follows Poisson distribution with λ varying from 1 to 6 (30 to 80).

Table 2.2: Proportion of Simulated Outbreaks being detected: Peak Pattern

Three Days Outbreak with a Peak						
Hyde County		λ_1				
		1	2	3	4	5
Fall,2006	Proposed Methods	100%	100%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	78%	86%	85%	97%	99%
	CUSUM	100%	100%	100%	100%	100%
Wake County		λ_1				
		5	10	15	20	25
Fall,2006	Proposed Methods	1%	6%	30%	59%	82%
	CUSUM	94%	100%	100%	100%	100%
Spring,2007	Proposed Methods	23%	29%	76%	92%	98%
	CUSUM	93%	100%	100%	100%	100%
Seven Days Outbreak with a Peak						
Hyde County		λ_1				
		1	2	3	4	5
Fall,2006	Proposed Methods	99%	100%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%
Spring,2007	Proposed Methods	88%	99%	100%	100%	100%
	CUSUM	100%	100%	100%	100%	100%
Wake County		λ_1				
		5	10	15	20	25
Fall,2006	Proposed Methods	1%	51%	100%	100%	100%
	CUSUM	83%	99%	100%	100%	100%
Spring,2007	Proposed Methods	70%	100%	100%	100%	100%
	CUSUM	81%	99%	100%	100%	100%

The start date of outbreak for fall, 2006 is November 10^{th}, and the start date of outbreak for spring, 2007 is April 10^{th}. Each outbreak last for 3 days. Simulated count in Hyde (Wake) follows Poisson distribution with starting/ending days intensity $\lambda_0 = 2$ ($\lambda_0 = 25$).

Table 2.3: Number of False Outbreaks Detected, Seven-Day Peak Pattern, Year 2006-2007

λ_1(Hyde) λ_1(Wake)	1 5	2 10	3 15	4 20	5 25
Proposed Method	137	136	134	128	127
CUSUM	2889	2889	2889	2889	2889

Fig. 2.1: Day-of-Week Effects in ED Asthma Admissions in North Carolina

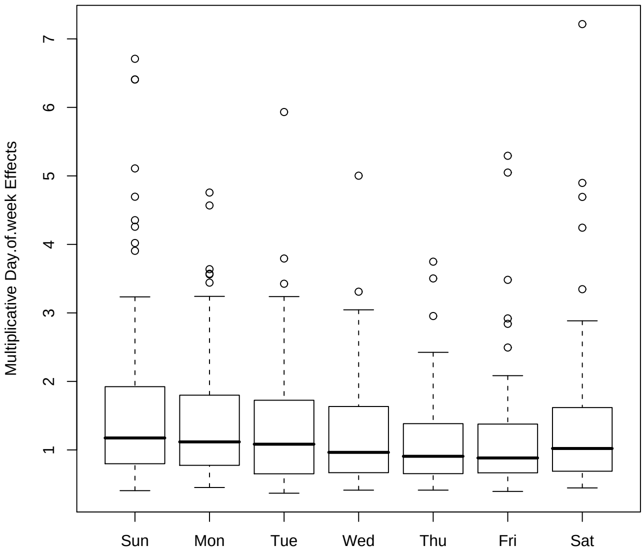


Figure 2.1: The average multiplicative day-of-week effects are: 1.66 on Sunday, 1.44 on Monday, 1.29 on Tuesday, 1.24 on Wednesday, 1.12 on Thursday, 1.16 on Friday and 1.33 on Saturday

Fig. 2.2: Q-Q Plot of Residuals Before and After Time Series Modeling for Wake County

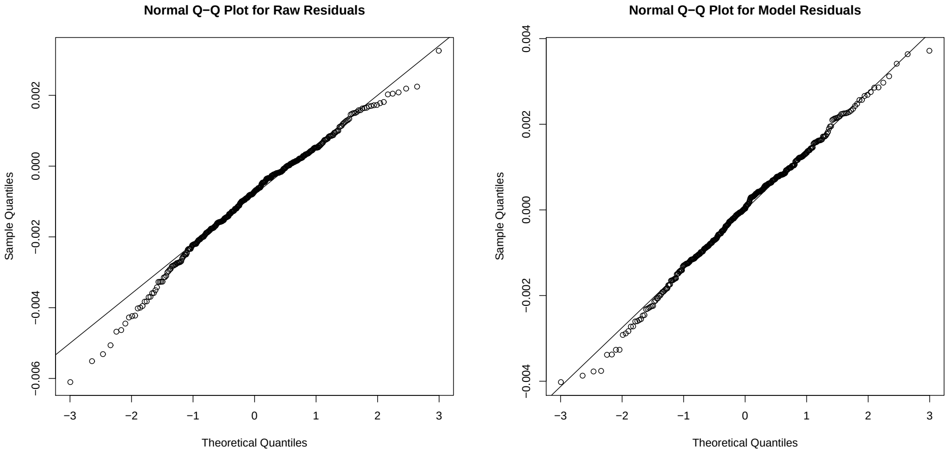


Fig. 2.3: Green shading shows how use of the proposed population distance metric affects the shape of North Carolina. In particular, the densely-populated areas in the center of the state are expanded, while the sparsely populated counties in Western North Carolina are shrunk. In this manner, a similar degree of person-level smoothing is applied regardless of population density.

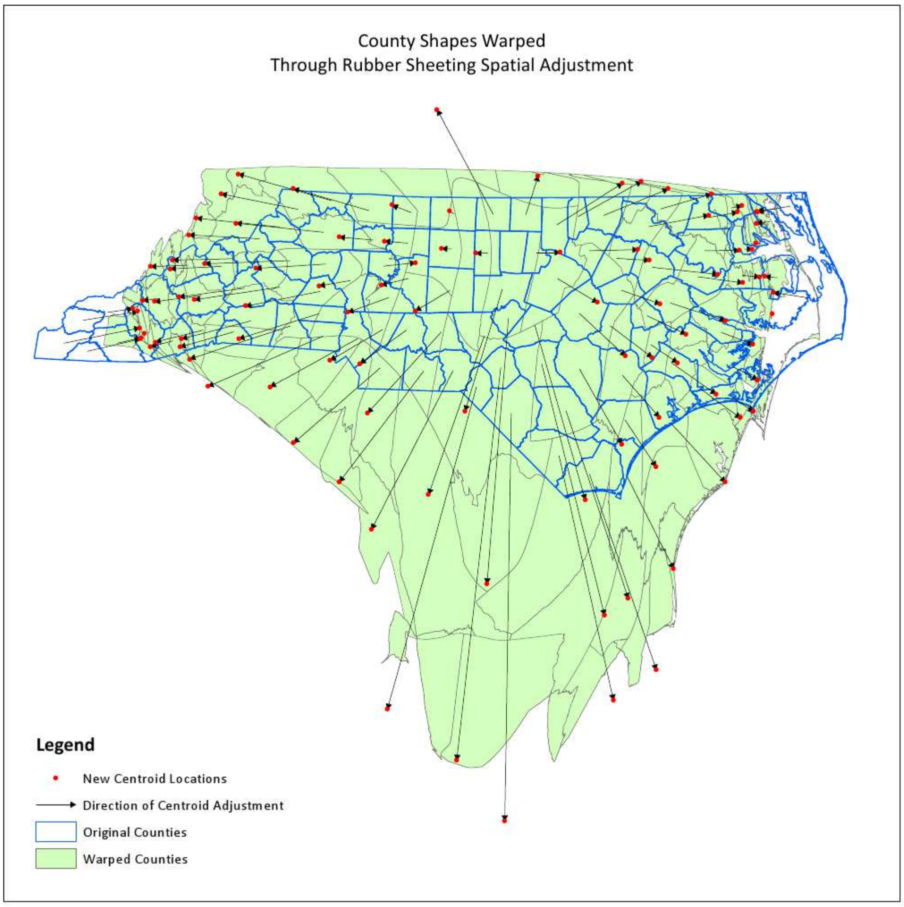


Fig. 2.4: Baseline Weekly Asthma Rate (per 100,000 population) of the 9^{th}, 22^{nd}, 35^{th} and 48^{th} Week in a Year

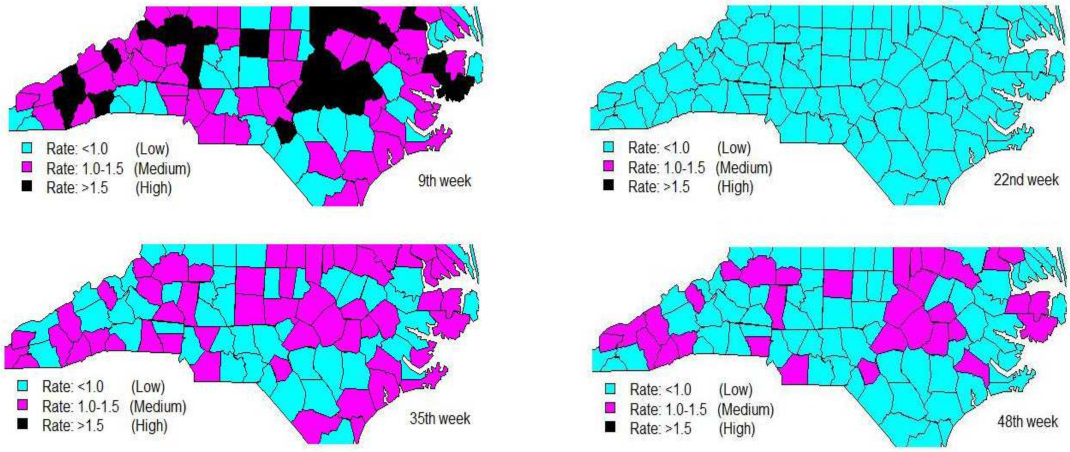
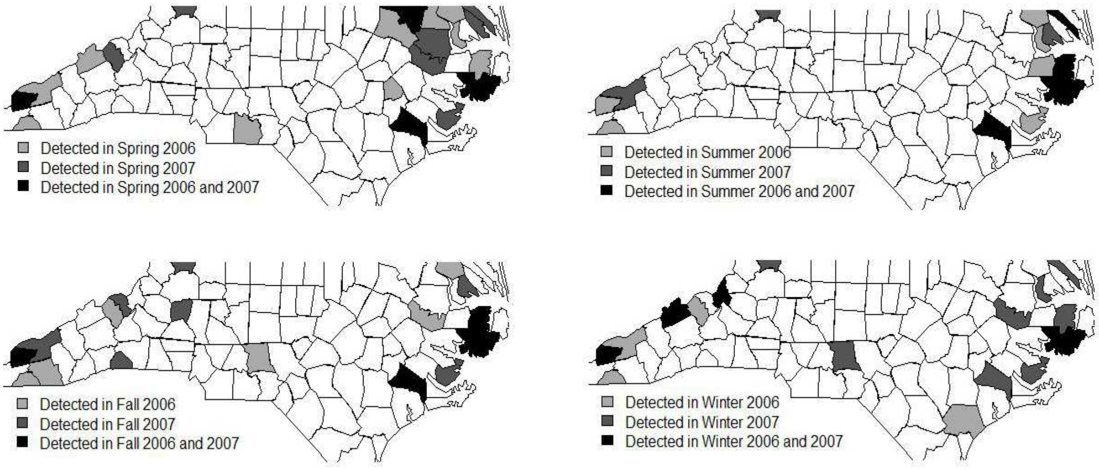


Fig. 2.5: Counties in North Carolina Detected for Potential Outbreaks



Chapter 3

Estimating Individualized Treatment Rules Using Outcome Weighted Learning

There is increasing interest in discovering individualized treatment rules for patients who have heterogeneous responses to treatment. In particular, one aims to find an optimal individualized treatment rule, which is a deterministic function of patient specific characteristics maximizing expected clinical outcome. In this chapter, it is shown that estimating such an optimal treatment rule is equivalent to a classification problem where each subject is weighted proportional to his or her clinical outcome. See Zhao et al. (2012). An outcome weighted learning approach is then proposed based on the support vector machine framework. We show that the resulting estimator of the treatment rule is consistent, and further obtain a finite sample bound for the difference between the expected outcome using the estimated individualized treatment rule and that of the optimal treatment rule. The performance of the proposed approach is demonstrated via simulation studies and an analysis of chronic depression data.

3.1 Introduction

In many different diseases, patients can show significant heterogeneity in response to treatments. One statistical approach for developing individual-adaptive interventions is to classify subjects into different risk levels estimated by a parametric or semiparametric regression model using prognostic factors, and then to assign therapy according to risk level (Eagle et al., 2004; Marlowe et al., 2007; Cai et al., 2010). However, the parametric or semiparametric model assumptions may not be valid due to the complexity of the disease mechanism and individual heterogeneity. Moreover, these approaches require preknowledge in allocating the optimal treatment to each risk category. There is also a significant literature examining discovery and development of personalized treatment relying on predicting patient responses to optional regimens (Rosenwald et al., 2002; van't Veer and Bernards, 2008), where the optimal decision leads to the best predicted outcome. One recent paper by Qian and Murphy (2011) applies a two-step procedure which first estimates a conditional mean for the response and then estimates the rule maximizing this conditional mean. A rich linear model is used to sufficiently approximate the conditional mean, with the estimated rule derived via l_1 penalized least squares (l_1-PLS). The method includes variable selection to facilitate parsimony and ease of interpretation. The conditional mean approximation requires estimating a prediction model of the relationship between pretreatment prognostic variables, treatments and clinical outcome using a prediction model. Reduction in the mean response is related to the excess prediction error, through which an upper bound can be constructed for the mean reduction of the associated treatment rule. However, by inverting the model to find the optimal treatment rule, this method emphasizes prediction accuracy of the clinical response model instead of directly optimizing the decision rule.

In this Chapter, we proposed a new method for solving this problem which circumvents the need for conditional mean modeling followed by inversion by directly estimating the decision rule which maximizes clinical response. Specifically, we demonstrate that the optimal treatment rule can be estimated within a weighted classification framework, where the weights are determined from the clinical outcomes. We then alleviate the computational problem by substituting the 0-1 loss in the classification with a convex surrogate loss as is done with the support vector machine (SVM) via the hinge loss (Cortes and Vapnik, 1995). The directness of this outcome weighted learning (OWL) approach enables us to better select targeted therapy while making full use of available information.

This chapter is organized as follows. In Section 3.2, we provide the mathematical concepts and framework for individualized treatment rules, and then formulate the problem as OWL. The proposed weighted SVM approach for constructing the optimal ITR is then developed in detail. In Section 3.3, consistency and risk bound results are established for the estimated rules. Faster convergence rates can be achieved with additional marginal assumptions on the data generating distribution. We present simulation studies to evaluate performance of the proposed method in Section 3.4. The method is then illustrated on the Nefazodone-CBASP data (Keller et al., 2000) in Section 3.5. We close this chapter with a short discussion in Section 3.6. The proofs of theoretical results are given in the Appendix 1.

3.2 Methodology

3.2.1 Individualized Treatment Rule (ITR)

We assume the data are collected from a two-arm randomized trial. That is, treatment assignments, denoted by $A \in \mathcal{A} = \{-1, 1\}$, are independent of any patient's prognostic variables, which are denoted as a d-dimensional vector $X = (X_1, \ldots, X_d)^T \in \mathcal{X}$. We let R be the observed clinical outcome, also called the "reward," and assume that

R is bounded, with larger values of R being more desirable. Thus an individualized treatment rule (ITR) is a map from the space of prognostic variables, $\mathcal{X}$, to the space of treatments, $\mathcal{A}$. An optimal ITR is a rule that maximizes the expected reward if implemented.

Mathematically, we can quantify the optimal ITR in terms of the relationship among (X, A, R). To see this, denote the distribution of (X, A, R) by P and expectation with respect to the P is denoted by E. For any given ITR $\mathcal{D}$, we let $P^{\mathcal{D}}$ denote the distribution of (X, A, R) given that $A = \mathcal{D}(X)$, i.e., the treatments are chosen according to the rule $\mathcal{D}$; correspondingly, the expectation with respect to $P^{\mathcal{D}}$ is denoted by $E^{\mathcal{D}}$. Then under the assumption that $P(A = a) > 0$ for $a = 1$ and -1, it is clear that $P^{\mathcal{D}}$ is absolutely continuous with respect to P and $dP^{\mathcal{D}}/dP = I(a = \mathcal{D}(x))/P(A = a)$, where $I(\cdot)$ is the indicator function. Thus, the expected reward under the ITR $\mathcal{D}$ is given as

$$E^{\mathcal{D}}(R) = \int R dP^{\mathcal{D}} = \int R \frac{dP^{\mathcal{D}}}{dP} dP = E\left[\frac{I(A = \mathcal{D}(X))}{A\pi + (1-A)/2} R\right],$$

where $\pi = P(A = 1)$. This expectation is called the value function associated with $\mathcal{D}$ and is denoted $\mathcal{V}(\mathcal{D})$. Consequently, an optimal ITR, $\mathcal{D}^*$, is a rule that maximizes $\mathcal{V}(\mathcal{D})$, i.e.,

$$\mathcal{D}^* \in \underset{\mathcal{D}}{\operatorname{argmax}} \, E\left[\frac{I(A = \mathcal{D}(X))}{A\pi + (1-A)/2} R\right].$$

Note that $\mathcal{D}^*$ does not change if R is replaced by $R + c$ for any constant c. Thus, without loss of generality, we assume that R is positive.

3.2.2 OWL for Estimating Optimal ITR

Assume that we observe i.i.d data $(X_i, A_i, R_i), i = 1, ..., n$ from the two-arm randomized trial described above. Previous approaches to estimating optimal ITR first estimate $E(R|X, A)$, using the observed data via parametric or semiparametric models, and then estimate the optimal decision rule by comparing the predicted value $E(R|X, A = 1)$ versus $E(R|X, A = -1)$ (Robins, 2004; Moodie et al., 2009; Qian and Murphy, 2011). As discussed before, these approaches indirectly estimate the optimal ITR, and are likely to produce a suboptimal ITR if the model for R given (X, A) is overfitted. As an alternative, we propose a nonparametric approach which directly maximizes the value function based on an outcome weighted learning method.

To illustrate our approach, we first notice that searching for the optimal ITR, $\mathcal{D}^*$, which maximizes $\mathcal{V}(\mathcal{D})$, is equivalent to finding $\mathcal{D}^*$ that minimizes

$$E[R|A = 1] + E[R|A = -1] - \mathcal{V}(\mathcal{D}) = E\left[\frac{I(A \neq \mathcal{D}(X))}{A\pi + (1 - A)/2} R\right].$$

The latter can be viewed as a weighted classification error, for which we want to classify A using X but we also weigh each misclassification event by $R/(A\pi + (1-A)/2)$. Hence, using the observed data, we approximate the weighted classification error by

$$n^{-1} \sum_{i=1}^{n} \frac{R_i}{A_i\pi + (1 - A)/2} I(A_i \neq \mathcal{D}(X_i))$$

and seek to minimize this expression to estimate $\mathcal{D}^*$. Since $\mathcal{D}(x)$ can always be represented as $\mathrm{sign}(f(x))$, for some decision function f, minimizing the above expression for $\mathcal{D}^*$ is equivalent to minimizing

$$\sum_{i=1}^{n} n^{-1} \frac{R_i}{A_i\pi + (1 - A)/2} I(A_i \neq \mathrm{sign}(f(X_i))) \tag{3.2.1}$$

to obtain the optimal f^*, and then setting $\mathcal{D}^*(x) = \mathrm{sign}(f^*(x))$.

The above minimization also has the following interpretation. That is, we intend to find a decision rule which assigns treatments to each subject only based on their prognostic information. For subjects observed to have a large reward, this rule is apt to recommend the same treatment assignments that the subject has actually received; however, for subjects with small rewards, the rule is more likely to give the opposite treatment assignment to what they received. In other words, if we stratify subjects into different strata based on the rewards, we will expect that the optimal ITR misclassifies less subjects in the high reward stratum as compared to the low reward stratum.

In the machine learning literature, (3.2.1) can be viewed as a weighted summation of 0-1 loss. It is well known that minimizing (3.2.1) is difficult due to the discontinuity and non-convexity of 0-1 loss. To alleviate this difficulty, one common approach is to find a convex surrogate loss for the 0-1 loss in (3.2.1) and develop a tractable estimation procedure (Zhang, 2004; Lugosi and Vayatis, 2004; Steinwart, 2005). Among many choices of surrogate loss, one of the most popular is the hinge loss used in the context of the support vector machine (Cortes and Vapnik, 1995), which we will adopt in this dissertation. Furthermore, we penalize the complexity of the decision function in order to avoid overfitting. In other words, instead of minimizing (3.2.1), we aim to minimize

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{A_i\pi+(1-A)/2}(1-A_i(f(X_i)))^{+}+\lambda_n\|f\|^2, \tag{3.2.2}$$

where $x^{+}=\max(x,0)$ and $\|f\|$ is some norm for f. In this way, we cast the problem of estimating the optimal ITR into a weighted classification problem using support vector machine techniques.

3.2.3 Linear Decision Rule for Optimal ITR

Suppose that the decision function $f(x)$ minimizing (3.2.2) is a linear function of x, that is, $f(x)=\langle\beta,x\rangle+\beta_0$, where $\langle\cdot,\cdot\rangle$ denotes the inner product in Euclidean space. Then the corresponding ITR will assign a subject with prognostic value X into treatment 1 if $\langle \beta, X \rangle + \beta_0 > 0$ and -1 otherwise.

In (3.2.2), we define $\|f\|$ as the Euclidean norm of β. Following the usual SVM, we introduce a slack variable ξ_i for subject i to allow a small portion of wrong classification. Denote $C > 0$ as the classifier margin. Then minimizing (3.2.2) can be rewritten as

$$\max_{\beta, \beta_0, \|\beta\|=1} C \text{ subject to } A_i(\langle \beta, X_i \rangle + \beta_0) \geq C(1 - \xi_i), \xi_i \geq 0, \sum \frac{R_i}{\pi_i} \xi_i < s,$$

where $\pi_i = \pi I(A_i = 1) + (1 - \pi) I(A_i = -1)$ and s is a constant depending on λ_n. This is equivalent to

$$\min \frac{1}{2}\|\beta\|^2 \text{ subject to } A_i(\langle \beta, X_i \rangle + \beta_0) \geq (1 - \xi_i), \xi_i \geq 0, \sum \frac{R_i}{\pi_i} \xi_i < s,$$

that is

$$\min \frac{1}{2}\|\beta\|^2 + \kappa \sum_{i=1}^{n} \frac{R_i}{\pi_i} \xi_i \quad \text{subject to} \quad A_i(\langle \beta, X_i \rangle + \beta_0) \geq (1 - \xi_i), \xi_i \geq 0,$$

where $\kappa > 0$ is a tuning parameter and R_i / π_i is the weight for the i^{th} point. We observe that the main difference compared to standard SVM is that we weigh each slack variable ξ_i with R_i / π_i.

After introducing Lagrange multipliers, the Lagrange function becomes:

$$\frac{1}{2}\|\beta\|^2 + \kappa \sum_{i=1}^{n} \frac{R_i}{\pi_i} \xi_i - \sum_{i=1}^{n} \alpha_i \{A_i(X_i^T \beta + \beta_0) - (1 - \xi_i)\} - \sum_{i=1}^{n} \mu_i \xi_i,$$

with $\alpha_i \geq 0, \mu_i \geq 0$. Taking derivatives with respect to (β, β_0) and ξ_i, we have $\beta = \sum_{i=1}^{n} \alpha_i A_i X_i, 0 = \sum_{i=1}^{n} \alpha_i A_i$ and $\alpha_i = \kappa R_i / \pi_i - \mu_i$. Plugging these equations into the

Lagrange function, we obtain the dual problem

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j A_i A_j \langle X_i, X_j \rangle$$

subject to $0 \leq \alpha_i \leq \kappa R_i / \pi_i, i = 1, \ldots, n$, and $\sum_{i=1}^{n} \alpha_i A_i = 0$. Quadratic programming algorithms from many widely available software packages can be used to solve this dual problem. Finally, we obtain that

$$\hat{\beta} = \sum_{\hat{\alpha}_i > 0} \hat{\alpha}_i A_i X_i,$$

and $\hat{\beta}_0$ can be solved using the margin points $(0 < \hat{\alpha}_i, \hat{\xi}_i = 0)$ subject to the Karush-Kuhn-Tucker conditions (Page 421, Hastie, Tibshirani & Friedman 2009). The decision rule is given by $\text{sign}\{\langle \hat{\beta}, X \rangle + \hat{\beta}_0\}$. Similar to the traditional SVM, the estimated decision rule is determined by the support vectors with $\hat{\alpha}_i > 0$.

3.2.4 Nonlinear Decision Rule for Optimal ITR

The previous section targets a linear boundary of prognostic variables. This may not be practically useful since the dimension of the prognostic variables can be quite high and complicated relationships may be involved between the desired treatments and these variables. However, we can easily generalize the previous approach to obtain a nonlinear decision rule for obtaining the optimal ITR.

We let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, called a kernel function, be continuous, symmetric and positive semidefinite. Given a real-valued kernel function k, we can associate with it a *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_k$, which is the completion of the linear span of all functions $\{k(\cdot, x), x \in \mathcal{X}\}$. The norm in $\mathcal{H}_k$, denoted by $\|\cdot\|_k$, is induced by the following inner product,

$$\langle f, g\rangle_k = \sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\beta_j k(x_i, x_j),$$

for $f(\cdot) = \sum_{i=1}^{n}\alpha_i k(\cdot, x_i)$ and $g(\cdot) = \sum_{j=1}^{m}\beta_j k(\cdot, x_j)$.

We note that our decision function $f(x)$ is from $\mathcal{H}_k$ equipped with norm $\|\cdot\|_k$. Thus since any function in $\mathcal{H}_k$ takes the form $\sum_{i=1}^{m}\alpha_i k(\cdot, x_i)$, it can be shown that the optimal decision function is given by

$$\sum_{i=1}^{n}\hat{\alpha}_i A_i k(X, X_i) + \hat{\beta}_0,$$

where $(\hat{\alpha}_1, ..., \hat{\alpha}_n)$ solves

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j A_i A_j k(X_i, X_j)$$

subject to $0 \leq \alpha_i \leq \kappa R_i/\pi_i, i = 1, \ldots, n$, and $\sum_{i=1}^{n}\alpha_i A_i = 0$. We note that if we choose $k(x, y) = \langle x, y\rangle$, then the obtained rule reduces to the previous linear rule.

3.3 Theoretical Results

In this section, we establish consistency of the optimal ITR estimated using OWL. We further obtain a risk bound for the estimated ITR and show how the bound can be improved for certain specific, realistic situations.

3.3.1 Notation

For any ITR $\mathcal{D}(x) = \text{sign}(f(x))$ associated with decision function $f(x)$, we define

$$\mathcal{R}(f) = E\left[\frac{R}{A\pi + (1-A)/2} I(A \neq \text{sign}(f(X)))\right]$$

and the minimal risk (called Bayes risk in the learning literature) as $\mathcal{R}^* = \inf_f\{\mathcal{R}(f)|f : \mathcal{X} \to \mathbb{R}\}$. Thus, for the optimal ITR $\mathcal{D}^*(x) = \text{sign}(f^*(x))$ (called the Bayes classifier in the learning literature), $\mathcal{R}^* = \mathcal{R}(f^*)$. In terms of the value function, we note that $\mathcal{V}(\mathcal{D}^*) - \mathcal{V}(\mathcal{D}) = \mathcal{R}(f) - \mathcal{R}(f^*)$.

In the OWL approach, we substitute 0-1 loss $I(A \neq \text{sign}(f(X)))$ by a surrogate loss, $\phi(Af(X))$, where $\phi(t) = (1-t)^+$. Thus we define the ϕ-risk

$$\mathcal{R}_\phi(f) = E\left[\frac{R}{A\pi + (1-A)/2}\phi(Af(X))\right],$$

and, similarly, the minimal $\phi-$risk as $\mathcal{R}^*_\phi = \inf_f\{\mathcal{R}_\phi(f)|f : \mathcal{X} \to \mathbb{R}\}$.

Recall that the estimated optimal ITR is given by $\text{sign}(\hat{f}_n(X))$, where

$$\hat{f}_n = \underset{f\in\mathcal{H}_k}{\text{argmin}}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i}\{1 - A_i f(X_i)\}^+ + \lambda_n\|f\|_k^2\right\}. \tag{3.3.1}$$

3.3.2 Fisher Consistency

We establish Fisher consistency of the decision function based on surrogate loss $\phi(t)$. Specifically, the following result holds:

Proposition 3.3.1. *If $\tilde{f}$ minimizes $R_\phi(f)$, then $\mathcal{D}^*(x) = sign(\tilde{f}(x))$.*

Proof. First, we note

$$\mathcal{D}^*(x) = \text{sign}\left\{E[R|X = x, A = 1] - E[R|X = x, A = -1]\right\}.$$

Next, for each $x \in \mathcal{X}$,

$$
\begin{aligned}
&E\left(R\frac{\phi(Af(X))}{A\pi + (1-A)/2}\Big| X = x\right)\\
=&E(R|A = 1, X = x)(1 - f(x)) + E(R|A = -1, X = x)(1 + f(x))\\
=&((E(R|A = -1, X = x) - E(R|A = 1, X = x))f(x)\\
+&E(R|A = -1, X = x) + E(R|A = 1, X = x)).
\end{aligned}
$$

Therefore, $\tilde{f}(x)$, which minimizes $\mathcal{R}_\phi(f)$, should be positive if $E(R|A = 1, X = x) > E(R|A = -1, X = x)$ and negative if $E(R|A = 1, X = x) < E(R|A = -1, X = x)$. That is, $\tilde{f}(x)$ has the same sign as $\mathcal{D}^*(x)$. The result holds.

This theorem justifies the validity of using $\phi(t)$ as the surrogate loss in OWL.

3.3.3 Excess Risk for $\mathcal{R}(f)$ and $\mathcal{R}_\phi(f)$

The following result shows that for any decision function f, the excess risk of f under 0-1 loss is no larger than the excess risk of f under the hinge loss. Thus, the loss of the value function due to the ITR associated with f can be bounded by the excess risk under the hinge loss. The proof of the theorem can be found in the Appendix 1.

Theorem 3.3.2. *For any measurable $f : \mathcal{X} \to \mathbb{R}$ and any probability distribution for (X, A, R),*

$$
\mathcal{R}(f) - \mathcal{R}^* \leq \mathcal{R}_\phi(f) - \mathcal{R}_\phi^*. \qquad (3.3.2)
$$

The proof follows the general arguments of Bartlett et al. (2006), in which they bound the risk associated with 0-1 loss in terms of the risk from surrogate loss, utilizing a convexified variational transform of the surrogate loss. In our proof, we extend this concept to our setting by establishing the validity of a weighted version of such a transformation.

3.3.4 Consistency and Risk Bounds

The purpose of this section is to establish the consistency of $\hat{f}_n$, and, moreover, to derive the convergence rate of $\mathcal{R}(\hat{f}_n) - \mathcal{R}^*$.

First, the following theorem shows that the risk due to $\hat{f}_n$ does converge to $\mathcal{R}^*$, and, equivalently, the value of $\hat{f}_n$ converges to the optimal value function. The proof of the theorem is deferred to the Appendix 1.

Theorem 3.3.3. *Assume that we choose a sequence $\lambda_n > 0$ such that $\lambda_n \to 0$ and $\lambda_n n \to \infty$. Then for all distributions P, we have that in probability,*

$$\lim_{n\to\infty} \left\{ \mathcal{R}_\phi(\hat{f}_n) - \inf_{f \in \bar{\mathcal{H}}_k} \mathcal{R}_\phi(f) \right\} = 0.$$

Thus, if f^* belongs to the closure of $\limsup_n \mathcal{H}_k$, where $\mathcal{H}_k$ depends on parameters varying with n, we have $\lim_{n\to\infty} \mathcal{R}_\phi(\hat{f}_n) = \mathcal{R}_\phi^*$ in probability. It then follows that $\lim_{n\to\infty} \mathcal{R}(\hat{f}_n) = \mathcal{R}^*$ in probability.

One special situation where f^* belongs to the limit space of $\mathcal{H}_k$ is when we choose $\mathcal{H}_k$ to be an RKHS with Gaussian kernel and let the kernel bandwidth decrease to zero. This will be shown in Theorem 3.3.4 below.

We now wish to derive the convergence rate of $\mathcal{R}(\hat{f}_n) - \mathcal{R}^*$ under certain regularity conditions on the distribution P. Specifically, we need the following "geometric noise" assumption for P (Steinwart and Scovel, 2007): Let

$$\eta(x) = \frac{E[R|X = x, A = 1] - E[R|X = x, A = -1]}{E[R|X = x, A = 1] + E[R|X = x, A = -1]} + 1/2, \tag{3.3.3}$$

then $2\eta(x) - 1$ is the decision boundary for the optimal ITR. We further define $\mathcal{X}^+ = \{x \in \mathcal{X} : 2\eta(x) - 1 > 0\}$, and $\mathcal{X}^- = \{x \in \mathcal{X} : 2\eta(x) - 1 < 0\}$. A distance function to the boundary between $\mathcal{X}^+$ and $\mathcal{X}^-$ is $\Delta(x) = \tilde{d}(x, \mathcal{X}^+)$ if $x \in \mathcal{X}^-$, $\Delta(x) = \tilde{d}(x, \mathcal{X}^-)$ if

$x \in \mathcal{X}^+$ and $\Delta(x) = 0$ otherwise, where $\tilde{d}(x, \mathcal{O})$ denotes the distance of x to a set $\mathcal{O}$ with respect to the Euclidean norm. Then the distribution P is said to have geometric noise exponent $0 < q < \infty$, if there exists a constant $C > 0$ such that

$$E\left[\exp\left(-\frac{\Delta(X)^2}{t}\right)|2\eta(X) - 1|\right] \leq Ct^{qd/2}, t > 0. \tag{3.3.4}$$

In some sense, this geometric noise exponent describes the behavior of the distribution in a neighborhood of the decision boundary. For example, for distinctly separable data, i.e., when $|2\eta(x) - 1| > \delta > 0$, for some constant δ, and η is continuous, q can be arbitrarily large.

In addition to this specific assumption for P, we also restrict the choice of RKHS to the space associated with Gaussian Radial Basis Function (RBF) kernels, i.e.,

$$k(x, x') = \exp(-\sigma_n^2\|x - x'\|^2), x, x' \in \mathcal{X},$$

where $\sigma_n > 0$ is a parameter varying with n. One advantage of using the Gaussian kernel is that we can determine the complexity of $\mathcal{H}_k$ in terms of capacity bounds with respect to the empirical L_2-norm, defined as

$$\|f - g\|_{L_2(P_n)} = \left(\frac{1}{n}\sum_{i=1}^{n}|f(X_i) - g(X_i)|^2\right)^{1/2}.$$

For any $\epsilon > 0$, the covering number of functional class $\mathcal{F}$ with respect to $L_2(P_n)$, $N(\mathcal{F}, \epsilon, L_2(P_n))$, is the smallest number of $L_2(P_n)$ ϵ-balls needed to cover $\mathcal{F}$, where an $L_2(P_n)$ ϵ-ball around a function $g \in \mathcal{F}$ is the set $\{f \in \mathcal{F} : \|f - g\|_{L_2(P_n)} < \epsilon\}$.

Specifically, according to Theorem 2.1 in Steinwart and Scovel (2007), we have that for any $\epsilon > 0$,

$$\sup_{P_n} \log N(B_{\mathcal{H}_k}, \epsilon, L_2(P_n)) \leq c_{\nu,\delta,d}\sigma_n^{(1-\nu/2)(1+\delta)d}\epsilon^{-\nu}, \tag{3.3.5}$$

where $B_{\mathcal{H}_k}$ is the closed unit ball of $\mathcal{H}_k$, and ν and δ are any numbers satisfying $0 < \nu \leq 2$ and $\delta > 0$.

Under the above conditions, we obtain the following theorem:

Theorem 3.3.4. *Let P be a distribution of (X, A, R) satisfying condition (4.3.5) with noise exponent $q > 0$. Then for any $\delta > 0, 0 < \nu < 2$, there exists a constant C (depending on ν, δ, d and π) such that for all $\tau \geq 1$ and $\sigma_n = \lambda_n^{-1/(q+1)d}$,*

$$Pr^*(\mathcal{R}(\hat{f}_n) \leq \mathcal{R}^* + \epsilon) \geq 1 - e^{-\tau},$$

where

$$\epsilon = C\left[\left(\frac{1}{\lambda_n}\right)^{\frac{2}{2+\nu}+\frac{(2-\nu)(1+\delta)}{(2+\nu)(1+q)}}\left(\frac{1}{n}\right)^{\frac{2}{2+\nu}} + \left(\frac{1}{\lambda_n}\right)^{\frac{q}{q+1}}\frac{\tau}{n} + \lambda_n^{\frac{q}{q+1}}\right].$$

The first two terms bound the stochastic error, which arises from the variability inherent in a finite sample size and which depends on the complexity of $\mathcal{H}_k$ in terms of covering numbers, while the third term controls the approximation error due to using $\mathcal{H}_k$, which depends on both σ_n and the noise behavior in the underlying distribution. We expect better approximation properties when the RKHS is more complex, but, conversely, we also expect larger stochastic variability. Using the above expression, an optimal choice of λ_n that balances bias and variance is given by

$$\lambda_n = (n)^{-\frac{2(1+q)}{(4+\nu)q+2+(2-\nu)(1+\delta)}},$$

so the optimal rate for the risk is

$$\mathcal{R}(\hat{f}_n) - \mathcal{R}^* = O_p\left((n)^{-\frac{2q}{(4+\nu)q+2+(2-\nu)(1+\delta)}}\right).$$

In particular, when data are well separated, q can be sufficiently large and we can let (δ, ν) be sufficiently small. Then the convergence rate almost achieves the "parametric" rate $n^{-1/2}$. However, if the marginal distribution of $\mathcal{X}$ has continuous density along the boundary, it can be calculated that $q = 2/d$. In this case, the convergence rate is approximately $n^{-2/(d+2)}$. Clearly, the speed of convergence is slower with larger dimension of the prognostic variable space.

To prove Theorem 3.3.4, we note that according to Theorem 3.3.2, it suffices to prove the result for the excess ϕ risk. We also use the fact that

$$\mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi^* = \mathcal{R}_\phi(\hat{f}_n) - \inf_{\mathcal{H}_k} \mathcal{R}_\phi(f) + \inf_{\mathcal{H}_k} \mathcal{R}_\phi(f) - \mathcal{R}_\phi^*.$$

We will then bound the first difference on the right-hand side using the empirical counterpart plus the stochastic variability due to the finite sample approximation. The latter can be controlled using large deviation results from empirical processes and some preliminary bound for $\|\hat{f}_n\|_k$. The second difference on the right-hand side will be bounded by using the approximation property of the RHKS and the geometric noise assumption of the underlying distribution P. The details are provided in the Appendix 1.

3.3.5 Improved Rate with Data Completely Separated

In this section, we show that a faster convergence rate can be obtained if the data are completely separated. We assume

(A1) $\forall x \in \mathcal{X}$, $|\eta(x) - 1/2| \geq \eta_0$, where $\eta(x)$ is defined in (3.3.3), and η is continuous.

(A2) $\forall x \in \mathcal{X}, \min(\eta(x), 1 - \eta(x)) \geq \eta_1$.

Assumption (A1) can be referred as a "low noise" condition equivalent to $|E(R|A = 1, X) - E(R|A = -1, X)| \geq \eta_0$. Thus, a jump of $\eta(x)$ at the level of 1/2 requires a gap between the rewards gained from treatment 1 and -1 on the same patient. This assumption is an adaptation of the noise condition used in classical SVM to obtain fast learning rates and it is essentially equivalent to one of the conditions in Blanchard et al. (2008).

Theorem 3.3.5. *Assume that (A1) and (A2) are satisfied. For any* $\nu \in (0, 1)$ *and* $q \in (0, \infty)$*, let* $\lambda_n = O(n^{-1/(\nu+1)})$ *and* $\sigma_n = \lambda_n^{-1/(q+1)d}$*. Then*

$$\mathcal{R}(\hat{f}_n) - \mathcal{R}^* = O_p\left(n^{-\frac{1}{\nu+1}\frac{q}{q+1}}\right).$$

We can let q go to ∞ and ν go to zero, and this theorem shows that the convergence rate for $\mathcal{R}(\hat{f}_n) - \mathcal{R}(f^*)$ is almost n^{-1}, a much faster rate compared to what was given in Theorem 3.4. This result is similar to results for SVM described in Tsybakov (2004); Steinwart and Scovel (2007); Blanchard et al. (2008).

To prove Theorem 3.3.5, we can rewrite the minimization problem in (3.3.1) as:

$$\min_{S \in \mathbb{R}^+} \left\{ \min_{f : \|f\|_k \le S} \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\pi_i} \{1 - A_i f(X_i)\}^+ + \lambda S^2 \right\}.$$

Thus the problem can be viewed in the model selection framework: a collection of models are balls in $\mathcal{H}_k$, and for each model, we solve the penalized empirical ϕ-risk minimization to obtain an estimator $\hat{f}_n$. We can utilize a result for model selection, presented in Theorem 4.3 of Blanchard et al. (2008), to choose the model which yields the minimal penalized empirical ϕ-risk among all the models. Proof details are provided in the Appendix 1.

3.4 Simulation Study

We have conducted extensive simulations to assess the small-sample performance of the proposed method. In these simulations, we generate 50-dimensional vectors of prognostic variables $X_1, \ldots, X_{50}$, consisting of independent $U[-1,1]$ variates. The treatment A is generated from $\{-1,1\}$ independently of X with $P(A=1)=1/2$. The response R is normally distributed with mean $Q_0 = 1 + 2X_1 + X_2 + 0.5X_3 + T_0(X,A)$ and standard deviation 1, where $T_0(X,A)$ reflects the interaction between treatment and prognostic variables and is chosen to vary according to the following four different scenarios:

1. $T_0(X,A) = 0.442(1 - X_1 - X_2)A$.

2. $T_0(X,A) = (X_2 - 0.25X_1^2 - 1)\,A$.

3. $T_0(X,A) = (0.5 - X_1^2 - X_2^2)\,(X_1^2 + X_2^2 - 0.3)\,A$.

4. $T_0(X,A) = (1 - X_1^3 + \exp(X_3^2 + X_5) + 0.6X_6 - (X_7 + X_8)^2)\,A$.

The decision boundaries in the first three scenarios are determined by X_1 and X_2. Scenario 1 corresponds to a linear decision boundary in the truth, where the shape of the boundary in Scenario 2 is a parabola. The third is a ring example, where the patients on the ring are assigned to one treatment, and another if inside or outside the ring. The decision boundary in the fourth example is fairly nonlinear in covariates, depending on covariates other than X_1 and X_2. For each scenario, we estimate the optimal ITR by applying OWL. We use the Gaussian kernel in the weighted SVM algorithm. There are two tuning parameters: λ_n, the parameter for penalty, and σ_n, the inverse bandwidth of the kernel. Since λ_n plays a role in controlling the severity of the penalty on the functions and σ_n determines the complexity of the function class utilized, σ_n should be chosen adaptively from the data simultaneously with λ_n. To illustrate this, Figure 3.1 shows the contours of the value function for the first scenario with different combinations of (λ_n, σ_n) when $n = 30$. We can see that λ_n interacts with σ_n, with larger λ_n generally coupled with smaller σ_n for equivalent value function levels. In our simulations, we apply a 5-fold cross validation procedure, in which we search over a pre-specified finite set of (λ_n, σ_n) to select the pair maximizing the average of the estimated values from the validation data. In case of tied values for parameter pair choices, we first choose the set of pairs with smallest λ_n and then select the one with largest σ_n.

Additionally, comparison is made among the following four methods:

- the proposed OWL using Gaussian kernel (OWL-Gaussian)
- the proposed OWL using linear kernel (OWL-Linear)
- the l_1 penalized least squares method (l_1-PLS) developed by Qian and Murphy (2011), which approximates $E(R|X, A)$ using the basis function set $(1, X, A, XA)$ and applies the LASSO method for variable selection, and
- ordinary least squares method (OLS), which estimates the conditional mean response using the same basis function set as in 3 but without variable selection.

We consider the OWL with linear kernel (method 2) mainly to assess the impact of different kernels in the weighted SVM algorithm. In this case, there is only one tuning parameter, λ_n, which can be chosen to maximize the value function in a cross-validation procedure. The selection of the tuning parameters in the l_1-PLS approach follows similarly. The last two approaches estimate the optimal ITR using the sign of the difference between the predicted $E(R|X, A = 1)$ and the predicted $E(R|X, A = -1)$. In the comparisons, the performances of the four methods are assessed by two criteria: the first criterion is to evaluate the value function using the estimated optimal ITR when applying to an independent and large validation data; the second criterion is to evaluate the misclassification rates of the estimated optimal ITR from the true optimal ITR using the validation data. Specifically, a validation set with 10000 observations is simulated to assess the performance of the approaches. The estimated value function using any ITR $\mathcal{D}$ is given by $\mathbb{P}_n^*[I(A = \mathcal{D}(X))R/P(A)]/\mathbb{P}_n^*[I(A = \mathcal{D}(X))/P(A)]$ (Murphy et al., 2001), where $\mathbb{P}_n^*$ denotes the empirical average using the validation data and $P(A)$ is the probability of being assigned treatment A.

For each scenario, we vary sample sizes for training datasets from 30 to 100, 200, 400 and 800, and repeat the simulation 1,000 times. The simulation results are presented in Figures 3.2 and 3.3, where we report the mean square errors (MSE) of both value functions and misclassification rates. Simulations show there are no large differences in the performance if we replace the Gaussian kernel with the linear kernel in the OWL. However, there are examples presenting advantages of the Gaussian kernel, which suggests that under certain circumstances, it is useful to have a flexible nonparametric estimation procedure to identify the optimal ITR for the underlying nonparametric structures. As demonstrated in Figure 3.2 and Figure 3.3, the OWL with either Gaussian kernel or linear kernel has better performance, especially for small samples, than the other two methods, from the points of view of producing larger value functions, smaller misclassification rates, and lower variability of the value function estimates. Specifically, when the approximation models used in the l_1-PLS and OLS are correct in the first scenario, the competing methods perform well with large sample size; however, the OWL still provides satisfactory results even if we use a Gaussian kernel. When the optimal ITR is nonlinear in X in the other scenarios, the OWL tends to give higher values and smaller misclassification rates. OLS generally fails unless the sample size is large enough since it encounters severe bias for small sample sizes. This is due to the fact that without variable selection for OLS, there is insufficient data to fit an accurate model with all 50 variables included. We also note that l_1-PLS has comparatively larger MSE, resulted from high variance of the method, which may be explained by the conflicting goals of maximizing the value function and minimizing the prediction error (Qian and Murphy, 2011).

3.5 Data Analysis

We apply the proposed method to analyze real data from the Nefazodone-CBASP clinical trial (Keller et al., 2000). The study randomized 681 outpatients with non-psychotic chronic major depressive disorder (MDD), in a 1:1:1 ratio to either Nafazodone, Cognitive Behavioral-Analysis System of Psychotherapy (CBASP) or the combination of Nefazodone and CBASP. The score on the 24-item Hamilton Rating Scale for Depression (HRSD) was the primary outcome, where higher scores indicate more severe depression. After excluding some patients with missing observations, we use a subset with 647 patients for analysis. Among them, 216, 220 and 211 patients were assigned to Nafazodone, CBASP and the combined treatment group respectively. Overall comparisons using t-tests show that the combination treatment had significant advantages over the other treatments with respect to HRSD scores obtained at end of the trial, while there are no significant differences between the nefazodone group and the psychotherapy group.

To estimate the optimal ITR, we perform pairwise comparisons between all combinations of two treatment arms, and, for each two-arm comparison, we apply the OWL approach. We only present the results from the Gaussian kernel, since the analysis shows a similarity with that of the linear kernel. Rewards used in the analyses are reversed HRSD scores and the prognostic variables X consist of 50 pretreatment variables. The results based on OWL are compared to results obtained using the l_1-PLS and OLS methods which use $(1, X, A, XA)$ in their regression models. For comparison between methods, we calculate the value function from a cross-validation type analysis. Specifically, the data is partitioned into 5 roughly equal-sized parts. We perform the analysis on 4 parts of the data, and obtain the estimated optimal ITRs using different methods. We then compute the estimated value functions using the remaining fifth part. The value functions calculated this way should better represent expected value functions for future subjects, as compared to calculating value functions based on the training data. The averages of the cross-validation value functions from the three methods are presented in Table 3.1.

From the table, we observe that OLS produces smaller value functions (corresponding to larger HRSD in the table) than the other two methods, possibly because of the high dimensional prognostic variable space. OWL performs similarly to l_1-PLS, but gives a 5% larger value function than l_1-PLS when comparing the Combination arm to the Nefazodone arm. In fact, when comparing combination treatment with nefazodone only, OWL recommends the combination treatment to all the patients in the validation data in each round of the cross validation procedure; the OLS assigns the combination treatment to around 70% of the patients in each validation subset; while the l_1-PLS recommends the combination to all the patients in three out of five validation sets, and 7% and 28% to the patients for the other two, indicating a very large variability. If we need to select treatment between combination and psychotherapy alone, the OWL approach recommends the combination treatment for all patients in the validation process. In contrast, the l_1-PLS chooses psychotherapy for 10 out of 86 patients in one round of validation, and recommends the combination for all patients in the other rounds. The percentages of patients who are recommended the combination treatment range from 66% to 85% across the five validation data sets when applying OLS. When the two single treatments are studied, there are only negligible differences in the estimated value functions from the three methods and the selection results also indicate an insignificant difference between them. In this case, about 20% of the patients are recommended to take the psycotherapy and the other 80% are recommended to be treated with nefazodone. Thus OWL not only yields ITRs with the best clinical outcomes, but the ITRs also have lowest variability compared to the other methods.

3.6 Discussion

The proposed OWL procedure appears to be more effective, across a broad range of possible forms of the interaction between prognostic variables and treatment, compared to previous methods. A two-stage procedure is likely to overfit the regression model, and thus cause troubles for value function approximation. The OWL provides a nonparametric approach which sidesteps the inversion of the predicted model required in other methods and benefits from directly maximizing the value function. The convergence rates for the OWL, aiming to identify the best ITR, nearly reach the optimal for the nonparametric SVM with the same type of assumptions on the separations. The rates, however, are not directly comparable to Qian and Murphy (2011), because we allow for complex multivariate interactions and formulate the problem in a nonparametric framework.

We only considered binary options for treatment. When there are more than two treatment classes, although we could do a series of pairwise comparisons as done in Section 3.5 above, this approach may not be optimal in terms of identifying the best rule considering all treatments simultaneously. It would thus be worthwhile to extend the OWL approach to settings involving three or more treatments. The case of multicategory SVM has been studied recently (Lee et al., 2004; Wang and Shen, 2006), and a similar generalization may be possible for finding ITRs involving three or more treatments. Another setting to consider is optimal ITR discovery for continuous treatments such as, for example, a continuous range of dose levels. In this situation, we could potentially utilize ideas underlying support vector regression (Vapnik, 1995), where the goal is to find a function that has at most ϵ deviation from the response. Using a similar rationale as the proposed OWL, we could develop corresponding procedures for continuous treatment spaces through weighing each subject by his/her clinical outcome.

Obtaining inference for individualized treatment regimens is also important and challenging. Due to high heterogeneities among individuals, there may be large variations in the estimated treatment rules across different training sets. Laber and Murphy (2011) construct an adaptive confidence interval for the test error under the non-regular framework. Confidence intervals for value functions help us determine whether essential differences exist among different decision rules. Thus an important future research topic is to derive the limiting distribution of $\mathcal{V}(\hat{\mathcal{D}}_n)-\mathcal{V}(\mathcal{D}^*)$ and to derive corresponding sample size formulas to aid in design of personalized medicine clinical trials.

In some complex diseases, dynamic treatment regimes may be more useful than the single-decision treatment rules studied in this chapter. Dynamic treatment regimes are customized sequential decision rules for individual patients which can adapt over time to an evolving illness. Recently, this research area has been of great interest in long term management of chronic disease. See, for example, Murphy et al. (2001); Thall et al. (2002); Murphy (2003); Robins (2004); Moodie et al. (2007); Zhao et al. (2011a). Extension of the proposed OWL approach to the dynamic setting would be of great interest and thus discussed in the next chapter.

Table 3.1: Mean Depression Scores (the Smaller, the Better) from Cross Validation Procedure with Different Methods

	OLS	l_1-PLS	OWL
Nefazodone vs CBASP	15.87	15.95	15.74
Combination vs Nefazodone	11.75	11.28	10.71
Combination vs CBASP	12.22	10.97	10.86

Fig. 3.1: Contour Plots of Value Function for Example 1 with $\lambda_n \in (0, 10)$ and $\sigma_n \in (0, 10)$

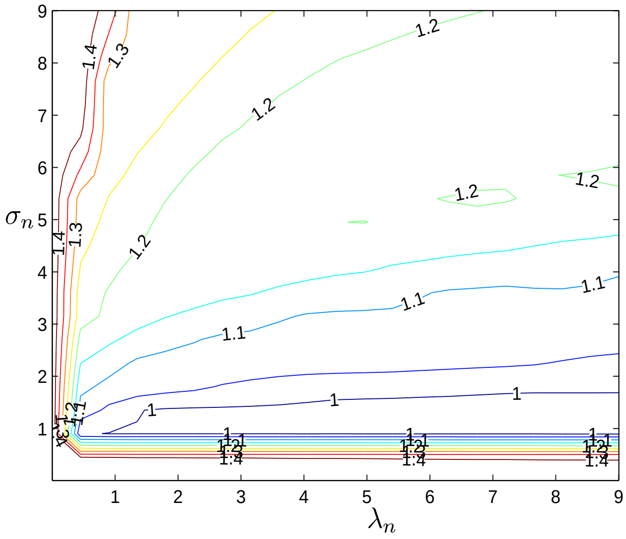


Fig. 3.2: MSE for Value Functions of Individualized Treatment Rules

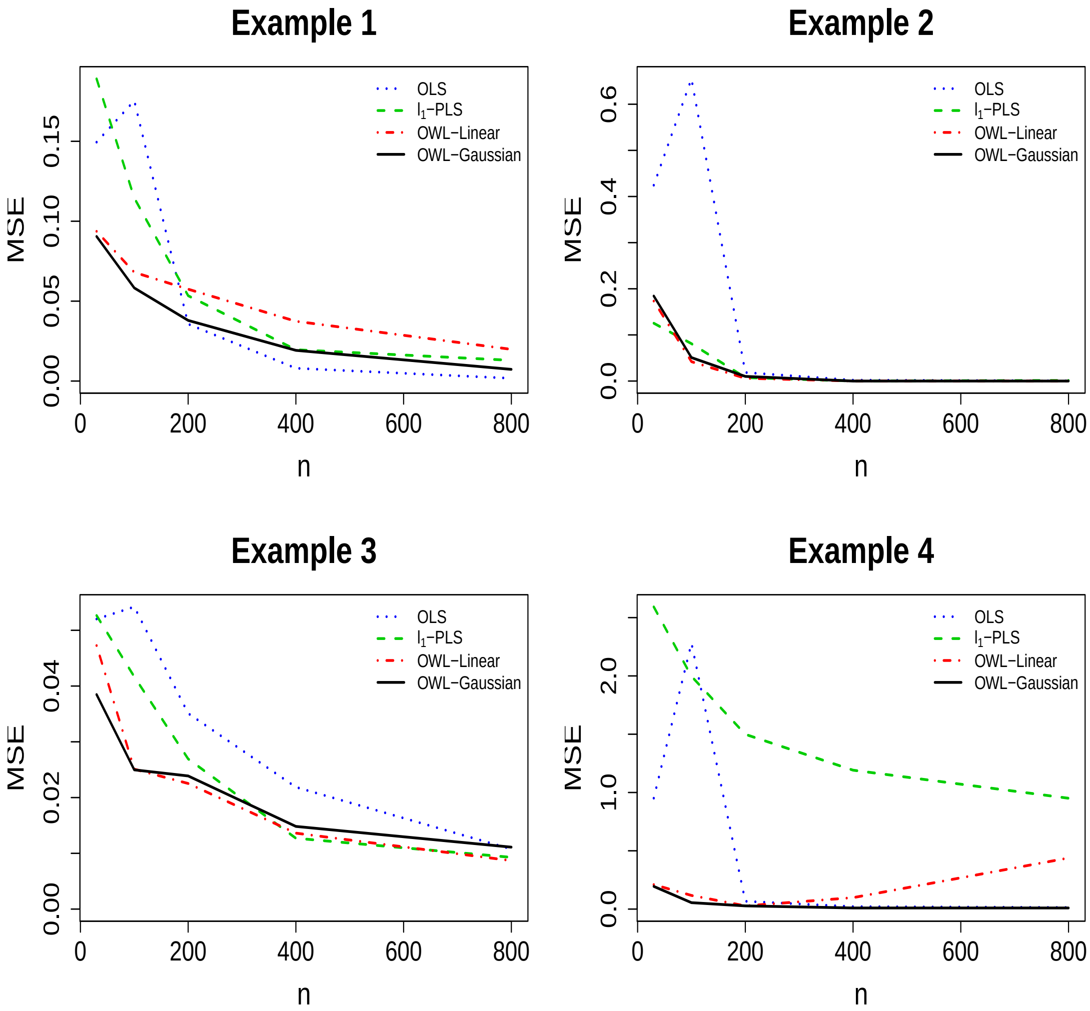
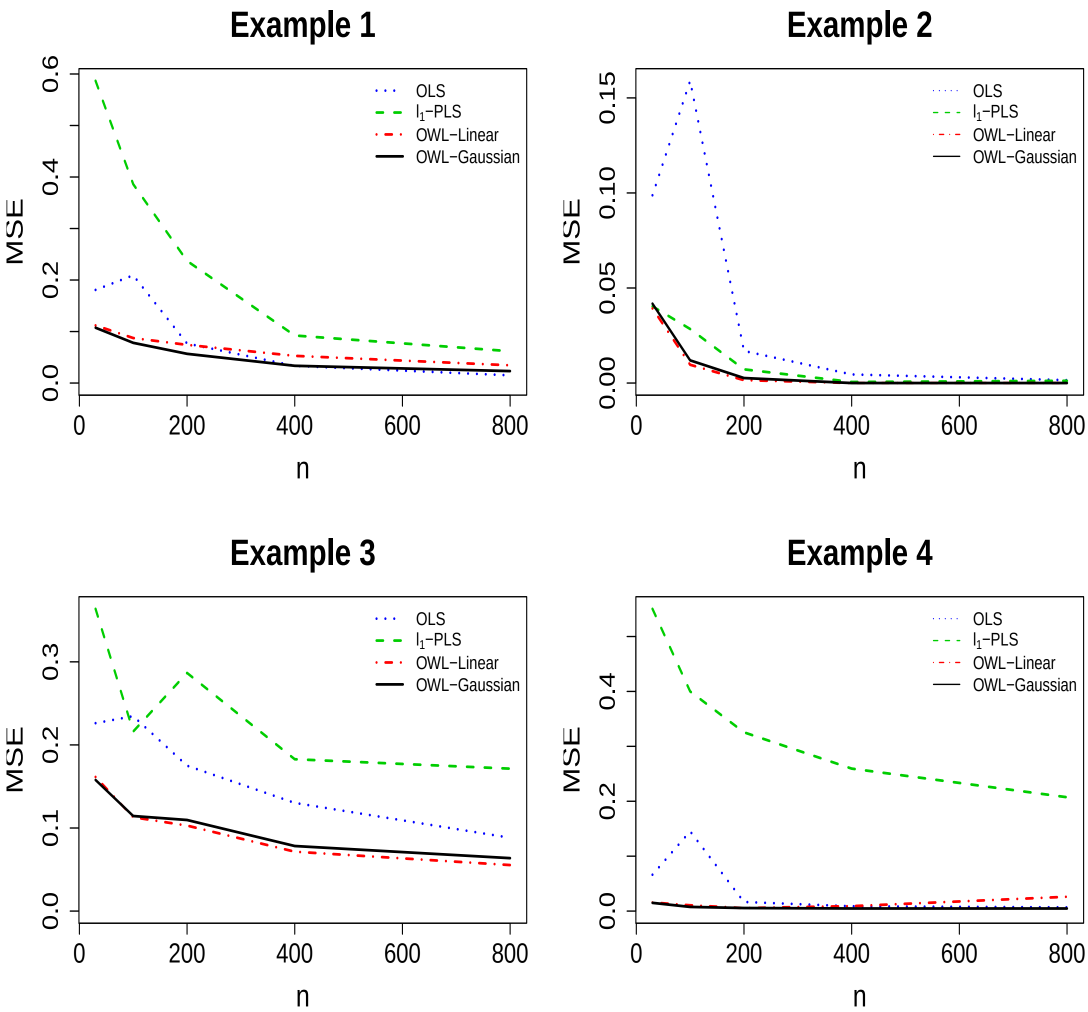


Fig. 3.3: MSE for Misclassification Rates of Individualized Treatment Rules



Chapter 4

Optimal Dynamic Treatment Regimes using Outcome Weighted Learning

In this chapter, we extend the previous proposed outcome weighted learning method to the multi-decision setup. The provided learning procedure is fundamentally different from Q-learning and other approaches currently used for finding optimal dynamic treatment regimes. Simulation results and a data analysis on the smoking cessation trial are presented to support the discussion.

4.1 Introduction

It is not uncommon that the optimal clinical strategies may require adaptation over time. Recognizing that there exist time varying characteristics among patients, and moreover, that the nature of diseases is as evolving and diversified as the people, clinicians have found out that even within the same patient, treatments which work now may not work later. This is especially common in the case of chronic diseases and conditions. Even if a “once and for all dosing” can be easily implemented, it may not be the best therapeutic plan for the patients’ health. For example, treatment for major depressive disorder is often driven by additional factors emerging over time, such as side-effect severity, treatment adherence and so on (Murphy et al., 2007); typically, the regimen for cancer patients involve multiple lines of treatment to prolong their lives or improve their progression free survival, such as non-small cell lung cancer (Socinski and Stinchcombe, 2007). Such problems have motivated plenty of literature on personalized treatment strategies, which ideally, should adapt with time dependent outcomes, including but not limited to patients response to previous treatments and side effects. Furthermore, rather than concentrating on a short-term benefit of a treatment, the goal here should be an improvement of the long-term gain by taking into account the treatment delayed effects to the patient.

Dynamic treatment regimes (DTR), also called adaptive treatment strategies (Murphy, 2005a), are sequential decision rules for individual patients which can adapt over time to an evolving illness. At each decision point, the covariate and treatment histories of a patient are taken as input for the decision rule, which outputs an individualized treatment recommendation subsequently. In general, we are interested in identifying the optimal dynamic treatment regime, defined as the sequence of decision rules that will maximize the mean response at the end of the time period.

A convenient way to formalize the problem in finding optimal dynamic treatment regimes is through potential outcomes (Neyman, 1990; Rubin, 1974, 1978; Robins, 1986), the value of the response variable that would be achieved, if contrary to the fact, the patient had been assigned to different treatments. Potential outcomes can be compared to find the regime that leads to the highest expected outcome if followed by the population. However, potential outcomes are not directly observable, since we can never observe all the results that could occur under different treatment regimes on the same patient. We need to construct estimators of the optimal dynamic treatment regimes using data from longitudinal studies.

Usually, several assumptions are made on the data:

- Stable Unit Treatment Value Assumption (SUTVA): A subject's outcome is unaffected by the treatment assignments to the other subjects (Rubin, 1986).
- Consistency: if the subject received the treatment assignments dictated by the regime, the observed outcomes will equal the potential outcomes under that regime (Robins, 1997).
- Positivity: the treatment patterns that match the dynamic treatment regime must have a positive probability of occurring. Thus, the information of treatment strategies are contained in the observed data to estimate their performance.
- No unmeasured confounders: the newly assigned treatments are independent of potential future outcomes from the treatment, conditional on the past and present observations.

There are different designs available to develop dynamic treatment regimes. A simple strategy is to randomize patients into possible groups with different treatment regimes at the baseline. However, lack of flexibilities and large sample size requirements may drive up the cost of such clinical trials. Alternatively, a sequential multiple assignment randomized trial (SMART) design has been advocated for this purpose(Lavori and Dawson, 2000, 2004; Dawson and Lavori, 2004; Murphy, 2005a; Murphy et al., 2007). In this design, multiple randomizations at the decision points are conducted, i.e., at each decision time, each patient is randomized to one of the possible treatments. The randomization probability may depend on all the observed information up to date. There are numerous SMART trials which have been conducted on different diseases, for example, prostate cancer (Thall et al., 2000), CATIE trial for Alzheimer disease (Schneider et al., 2001), STAR*D for depression (Rush et al., 2004), and a smoking cessation trial (Strecher et al., 2008). SMART designs guarantee that the assumption of no unmeasured confounders is satisfied.

Traditionally, people use dynamic programming to find the optimal decision rules, which requires a knowledge of the distribution of the entire process. Lavori and Dawson (2000) use multiple imputation techniques to estimate all potential outcomes, and the adaptive strategies can be compared based on the imputed outcomes. Murphy et al. (2001) employed a structural model to estimate the mean response that would have been observed if the whole population followed a particular dynamic treatment regime. Likelihood-based approaches are proposed in Thall et al. (2000, 2002, 2007), where both frequentist and Bayesian methods are applied to estimate parameters and thus the optimal strategies. Semiparametric methods are proposed to evaluate and compare different treatment policies when survival distributions are of interest in two-stage oncology trials (Lunceford et al., 2002; Wahed and Tsiatis, 2004, 2006). Two common approaches to constructing a dynamic treatment regime from data include Q-learning (Watkins, 1989; Sutton and Barto, 1998) and A-learning (Murphy, 2003; Blatt et al., 2004), where 'Q' denotes 'quality' and 'A' denotes 'advantage'. Q-Learning, originally proposed in the computer science literature, has become a powerful tool to discover optimal regimes in the clinical research arena (Murphy et al., 2007; Zhao et al., 2009, 2011a). Q-learning constructs decision rules through fitting Q-functions, which are the conditional mean functions of outcomes given the histories and treatments. When the dimension of the potential actions is small, linear regression methods should be adequate for fitting Q-functions, but in more extreme cases these methods can be problematic. A richer class of basis functions may be desirable for estimation, such as quadratic regression. The unclear and potentially complex structure of Q-functions has motivated researchers to seek other methods with more flexibility. Zhao et al.

(2009) utilized flexible nonparametric methods, specifically, the support vector machine and extremely randomized trees for fitting Q-functions. Murphy (2003) proposed a semiparametric method for estimating optimal decisions, where the regret function is modeled. Later Robins (2004) deduced optimal decision rules using structural nested mean models. Moodie et al. (2007) provided a nice discussion showing that the two methods are closely related, that Robins modeled effects relative to the predictor value of 0, and Murphy's model focused on the effects relative to the optimal predictor value. It turns out that Murphy's model is a special case of Robins'. Chakraborty et al. (2009) showed that Q-learning is an efficient version of Robins' method under certain conditions. For more discussion on the relationship between Q- and A-learning, we refer readers to Schulte et al. (2012). However, the optimal dynamic treatment regimes can not be obtained if models are misspecified or poorly fitted. Moreover, while the goal is to maximize the long term outcome, estimation based on minimizing the prediction error may not necessarily yield the desired results for decision making.

In this chapter, we provide novel methodologies for finding optimal dynamic treatment regimes, concentrating on directly maximizing the expected long term outcome without positing the model for the outcomes giving the patients' history information at each stage. For the single decision setup, we have shown that the optimization can be achieved within a weighted classification framework, where weights are determined by the clinical outcomes. A heuristic inspired by dynamic programming guides us to identify a sequence of optimal decision rules using the same algorithm but through a backwards recursive fashion. We also develop an outcome weighted learning procedure for two stage problems, where the computational burden arising from the two dimensional 0-1 losses is mitigated by using a two-dimensional surrogate loss function. The integration of statistical machine learning techniques and optimal decision rule discoveries has provided us alternative views on the problem.

The remainder of the chapter is organized as follows. In Section 4.2, we formulate the problem of finding the optimal dynamic treatment regimes in a mathematical framework. Two different approaches are proposed stemming from the concept of outcome weighted learning proposed in Chapter 3, while now taking into account the dynamic aspects. Section 4.3 provides theoretical justifications for the proposed methods in Section 4.2 and shows nice properties for finding the optimal DTR. We present an empirical comparison on the performances of the proposed methods and Q-learning. Section 4.5 focuses on the application of the proposed outcome weighted learning methods for the multi-decision setup, where the data comes from a smoking cessation trial. We provide a brief discussion in Section 4.6.

4.2 General Methodology

4.2.1 Dynamic Treatment Regimes (DTR)

Consider a multistage decision problem where decisions are made at set times $t \in \{1, \ldots, T\}$ with total sample size n. We use X and A to denote random variables, where X_t is the observation available at the t^{th} stage, and A_t is the treatment chosen at the t^{th} stage subsequent to observing X_t. We assume that A_t is binary with values taken in $\mathcal{A}_t = \{-1, 1\}$. Let $H_t = \{X_1, A_1, \ldots, X_{t-1}, A_{t-1}, X_t\} \in \mathcal{H}_t$ denote the covariate and treatment history available at the t^{th} stage, with the dimension of H_t denoted by p_t, $t = 1, \ldots, T$. In this dissertation, we consider a sequential multiple assignment randomized trial design, and assume that two possible treatments are randomized with known probabilities possibly dependent on H_t at each stage. Without loss of generality, we let $P(A_t = 1|H_t) = \pi_t, t = 1, \ldots, T$, where the π_ts are known constants. Corresponding lower cases are used to denote realizations of the random variables. Let the observed clinical outcome, also called the "reward", following the t^{th} stage be given by $R_t =$

$L(H_{t+1})$, where L is a deterministic function of all histories up to stage $t+1$. The R_t are assumed to be bounded, with larger values being more preferable. We define a dynamic treatment regime $\mathbf{d} \in \mathcal{D}$, as a sequence of deterministic decision rules, $\{d_1, \ldots, d_T\}$, mapping from the history space $\mathcal{H}_t$, to the space of available treatments $\mathcal{A}_t$ at the t^{th} stage. Given h_t, $d_t(h_t) = a_t$ is a treatment at stage t depending on history. The goal at each stage is to decide on the treatment which will lead to the maximized long-term benefit, that is, to find a DTR to optimize the overall mean outcome through the final follow-up time T.

Assume that the finite longitudinal trajectories are drawn at random from a fixed and unknown distribution P and denote expectations with respect to P by E, where a trajectory is defined as a realization of $(X_1, A_1, R_1, \ldots X_T, A_T, R_T)$. Let $P_{\mathbf{d}}$ denote the distribution of $(X_1, A_1, R_1, \ldots X_T, A_T, R_T)$ when regime $\mathbf{d}$ is used to assign treatments, and correspondingly, the expectation with respect to $P_{\mathbf{d}}$ is denoted by $E_{\mathbf{d}}$. We only consider the collection of all regimes, still denoted by $\mathcal{D}$, satisfying

$$P\left(\prod_{t=1}^{T} P(A_t = d_t(H_t)|H_t) > 0\right) = 1,$$

which indicates that treatments following regime $\mathbf{d} = \{d_1, \ldots, d_T\}$ can occur in the longitudinal data. Thus, the information of treatment strategies $\mathbf{d}$ are contained in the observed data to estimate their performance. To choose the regime that yields the most desirable long term consequence, we seek a DTR that maximizes the expectations of the sum of the rewards over the time trajectories. For $\mathbf{d} \in \mathcal{D}$, we establish the value function as,

$$V(\mathbf{d}) = E_{\mathbf{d}}\left[\sum_{t=1}^{T} R_t\right].$$

The optimal value function is defined as $V^* = \max_{\mathbf{d} \in \mathcal{D}} V(\mathbf{d})$, and the optimal dynamic treatment regime, denoted by $\mathbf{d}^*$, is the regime leading to the value function with the highest value. Additionally, the expected total reward that can be accumulated over the future from the t^{th} stage is defined as

$$V_t(h_t) = E_{\mathbf{d}}\left[\sum_{l=t}^{T} R_l | H_t = h_t\right],$$

and the optimal value function from stage t is $V_t^*(h_t) = \max_{\mathbf{d}\in\mathcal{D}} V_t(h_t)$.

A fundamental property of value functions is that they satisfy particular recursive relationships such as the Bellman equation (Bellman, 1957; Sutton and Barto, 1998). According to the Bellman optimality equation, we have

$$V_t^*(h_t) = \max_{a_t} E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = a_t), \tag{4.2.1}$$

with $V_{T+1}^*(H_{T+1}) = 0$. And the optimal regime for the t^{th} stage satisfies the following relation:

$$d_t^*(h_t) = \underset{a_t}{\operatorname{argmax}}\, E[R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = a_t],$$

where $V_{t+1}^*(H_{t+1})$ is the cumulative sum of rewards from stage $t+1$ to stage T when using the optimal regime $\mathbf{d}^*$ thereafter, given the history H_{t+1}.

Traditionally, people use dynamic programming to find the optimal decision rules. In this case, the complete probability distribution must be specified. Another popular method to construct dynamic treatment regime is Q-learning, which requires less computation. The time-dependent Q-function is defined as

$$Q_t(h_t, a_t) = E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = a_t).$$

Hence, there exists a recursive form as follows:

$$Q_t(h_t, a_t) = E(R_t + \max_{a_{t+1}} Q_{t+1}(H_{t+1}, a_{t+1})|H_t = h_t, A_t = a_t).$$

The optimal sequence of decision rules can be determined from $d_t^*(h_t) = \operatorname{argmax}_{a_t} Q_t(h_t, a_t)$. Usually, linear models are used for estimating Q-functions, and the optimal DTR can be identified given the history of the patient. However, Murphy (2005b) pointed out that there is a mismatch between Q-learning and the goal of learning a regime that maximizes the value function. In addition, the approximation space may not be complex enough to capture the Q-function structure, which can yield a large bias. In the section below, we propose approaches which enable us to directly estimate the dynamic treatment regime maximizing the value function.

4.2.2 Outcome Weighted Learning for the Multi-Decision Setup

In this section, we introduce our methods for constructing optimal dynamic treatment regimes from a sequential multiple assignment randomized trial. The developed methods identify the optimal strategy by directly maximizing the expected long term outcome (the value function). We have introduced the OWL method for the single stage decision problem, formulated in a weighted support vector machine framework in Chapter 3. This section is organized as follows. We first extend the methodology to the multiple stage decision problem by repeatedly estimating the optimal regimes backwards from the end of the study. We then provide an iterative procedure to find the optimal dynamic treatment regimes for a two-stage decision problem.

Backwards Outcome Weighted Learning (BOWL)

Assume that we observe a training data set $(X_{1i}, A_{1i}, R_{1i}, \ldots, X_{Ti}, A_{Ti}, R_{Ti}), i = 1, \ldots, n$ consisting of n i.i.d. patient trajectories from a SMART study. Outcome weighted learning can be applied in a backward fashion to yield the optimal decision for stage $t, t = 1, \ldots, T$. At the t^{th} stage, we can write the optimal value function

(4.2.1) in the following way that

$$V_t^*(h_t) = \max_{d_t} E\left[\left.\frac{(R_t + V_{t+1}^*(H_{t+1}))I(A_t = d_t(h_t))}{A_t\pi_t + (1-A_t)/2}\right| H_t = h_t\right].$$

Therefore, the optimal treatment rule d_t^* should minimize the risk function at that stage, defined as

$$\mathcal{R}_t(h_t) = E\left[\left.\frac{(R_t + V_{t+1}^*(H_{t+1}))I(A_t \neq \text{sign}(f_t(h_t)))}{A_t\pi_t + (1-A_t)/2}\right| H_t = h_t\right],$$

where $d_t(h_t)$ is represented as $\text{sign}(f_t(h_t))$ with some decision function f_t. Note that the risk function at the t^{th} stage is defined so that the optimal treatment regime d^* is applied thereafter.

For estimation purposes, we generally replace the expected value with its empirical analog in terms of the observed data, and then conduct the optimization. However, the resulting problem is difficult to optimize directly because of the discontinuity and non-convexity of the 0-1 loss. Similarly as in the one-decision point problem, we can mitigate the computational burden and develop a tractable estimation procedure by using a convex surrogate loss, such as hinge loss, in place of the 0-1 loss. In addition, we penalize the complexity of the decision function f_t to avoid overfitting. Hence, we aim to minimize

$$\mathbb{E}_n\left[\left.\frac{(R_t + V_{t+1}^*(H_{t+1}))\phi(A_t f_t(h_t))}{A_t\pi_t + (1-A_t)/2}\right| H_t = h_t\right] + \lambda_{t,n}\|f_t\|^2, \tag{4.2.2}$$

where $\mathbb{E}_n$ denotes the empirical measure of the observed data and $\lambda_{t,n}$ is the penalization parameter for stage t. We still need to find an estimate of $V_{t+1}^*(H_{t+1})$. Given that the optimal dynamic treatment regimes from stage $t+1$ to stage T have been estimated, one finds a natural estimate to be the mean response of all patients whose assigned treatments after stage $t+1$ are consistent with the estimated regime. Let $\hat{d}_{t,n}(h_t) = \text{sign}(\hat{f}_{t,n}(h_t)), t = 1, \ldots, T$ be the estimated rule at stage t, then

$$\hat{V}_{t+1,n}(\hat{f}_{t+1,n}, \ldots, \hat{f}_{T,n}) = \mathbb{E}_n \left(\sum_{l=t+1}^{T} R_l | A_l = \text{sign}(\hat{f}_{l,n}(H_l)), l = t+1, \ldots, T \right)$$

is the estimated value from stage $t+1$ to stage T when using the estimated regime $\text{sign}(\hat{f}_n)$ thereafter.

To fix ideas, we start BOWL with the final stage T, where $\hat{V}_{T+1,n} = 0$. Therefore, solving (4.2.2) reduces to a single stage problem. $\hat{d}_{T,n}(H_T)$ can be estimated following the developed weighted support vector machine procedure in Zhao et al. (2012) and $\hat{V}_{T,n}$ is obtained subsequently. We can then repeatedly find the optimal decision rule at each decision point by solving

$$\hat{f}_{t,n}(h_t) = \underset{f_t}{\text{argmin}} \left\{ \mathbb{E}_n \left[\frac{(R_t + \hat{V}_{t+1,n}(\hat{f}_{t+1,n}, \ldots, \hat{f}_{T,n}))\phi(A_t f_t(h_t))}{A_t \pi_t + (1 - A_t)/2} \right| H_t = h_t, \right.$$
$$\left. A_l = \text{sign}(\hat{f}_{l,n}(H_l)), l = t+1, \ldots, T \Big] + \lambda_{t,n} \|f_t\|^2 \right\}.$$

Consequently, to estimate the optimal decision rule for stage t, the analysis is restricted to the subset of patients who have been assigned to the estimated optimal treatments after that stage. The size of available data for estimation in the current step is decreased while implementing backwards outcome weighted learning, compared to the previous step. We denote the sample size for estimation at stage t by n_t. For example, under pure randomization with randomization probability 0.5 at each decision point, the size is reduced by half as we proceed compared to the previous estimation step, i.e., $n_t = n_{t+1}/2$.

If the decision function f_t at stage t is a linear function of h_t, $f_t(h_t) = \beta_t h_t + \beta_{0t}$, then $\|f_t\|$ is defined as the Euclidean norm of β_t. Consider a nonlinear decision rule, where f_t resides in a reproducing kernel hilbert space (RKHS) $\mathcal{H}_{k_t}$, associated with a real-valued kernel function k_t. The use of the kernel function enables us to deal with data of a more complex structure. In this case, $\mathcal{H}_{k_t}$ is equipped with the RKHS norm $\|\cdot\|_{k_t}$, and $\|f_t\|$ is defined as $\|f_t\|_{k_t}$.

Iterative Outcome Weighted Learning (IOWL)

Though the performance of backwards outcome weighted learning is robust under various sample sizes, as shown in later sections, we develop an iterative procedure for the two-stage setup to compensate for the fact that we do not make full use of the data across all stages. Specifically, the the objective is to find the dynamic treatment regime with a sequence of two decision rules, which can maximize $E_{\mathbf{d}}(R_1 + R_2)$, the expected total amount of reward when the treatments are chosen according to regime $\mathbf{d}$. For any DTR $(d_1(h_1), d_2(h_2)) = \text{sign}(f_1(h_1), f_2(h_2))$ associated with decision functions $f_1(h_1)$ and $f_2(h_2)$ in two stages respectively, the objective value function for maximizing the expected long term outcome, defined as $V(f_1, f_2)$, can be written in the following form,

$$V(f_1, f_2) = E\left[(R_1 + R_2)\frac{I(A_1 = \text{sign}(f_1(H_1)))I(A_2 = \text{sign}(f_2(H_2)))}{(A_1\pi_1 + (1 - A_1)/2)(A_2\pi_2 + (1 - A_2)/2)}\right]. \qquad (4.2.3)$$

The optimal dynamic treatment regime $\mathbf{d}^*$ leads to the maximal value, that is, $\mathbf{d}^* = (d_1^*, d_2^*) = \text{sign}(f_1^*, f_2^*)$, where $(f_1^*, f_2^*) = \text{argmax}_{f_1, f_2} V(f_1, f_2)$. And the optimal value as $V^* = V(f_1^*, f_2^*)$.

As stated in the previous section, after completing the analysis for stage 2, we only use a subset of data to estimate the optimal treatment rule for stage 1. A modification of the BOWL procedure can be implemented which potentially utilizes the whole data set to find (d_1^*, d_2^*). Upon obtaining the stage 1 estimated rule $\hat{d}_{1,n}$ using BOWL, we reestimate the optimal stage 2 rule $\hat{d}_{2,n}^{new}$ based on the subset of patients whose stage 1 treatment assignments are consistent with $\hat{d}_{1,n}$. We continue with the reestimation of the optimal stage 1 rule $\hat{d}_{1,n}^{new}$ using the information of patients with $A_2 = \hat{d}_{2,n}^{new}$. The process is then iterated until the estimated value converges. The iteration procedure is described as follows.

1. Estimate the optimal dynamic treatment regime $\text{sign}(\hat{f}_{1,n}, \hat{f}_{2,n})$ using BOWL.

2. Given $\hat{f}_{1,n}$, we find an updated optimal stage 2 treatment decision function $\hat{f}_{2,n}^{new}$ by maximizing

$$\mathbb{E}_n\left[\mathbb{E}_n\left((R_1+R_2)\frac{I(A_2=\text{sign}(f_2(H_2)))}{A_2\pi_2+(1-A_2)/2}\middle| A_1=\text{sign}(\hat{f}_{1,n}(H_1))\right)\right].$$

   Set $\hat{f}_{2,n} = \hat{f}_{2,n}^{new}$.

3. Substituting $\hat{f}_{2,n}$ into the value function, we obtain $\hat{f}_{1,n}^{new}$ by maximizing

$$\mathbb{E}_n\left[\mathbb{E}_n\left((R_1+R_2)\frac{I(A_1=\text{sign}(f_1(H_1)))}{A_1\pi_1+(1-A_1)/2}\middle| A_2=\text{sign}(\hat{f}_{2,n}(H_1))\right)\right].$$

   Set $\hat{f}_{1,n} = \hat{f}_{1,n}^{new}$.

4. We iterate between Step 2 and 3 until

$$|V(\hat{f}_{1,n}, \hat{f}_{2,n}) - V(\hat{f}_{1,n}^{new}, \hat{f}_{2,n}^{new})| \leq \epsilon$$

   for a prespecified threshold ϵ.

In each step, we only update the decision rule for one stage while leaving the other unchanged. The value gets replaced by a hopefully better estimate after each iteration, which can be elaborated by calculating the form of the decision rule. In step 2, the updated rule can be obtained as a function of $\hat{f}_{1,n}$, denoted as $\hat{f}_{2,n}^{new} = T_2(\hat{f}_{1,n})$, where

$$\text{sign}(\hat{f}_{2,n}^{new}) = \text{sign}[E(R_2|A_1 = \text{sign}(\hat{f}_{1,n}), X_2, A_2 = 1) - E(R_2|A_1 = \text{sign}(\hat{f}_{1,n}), X_2, A_2 = -1)].$$

Letting $Q_2^*(\hat{f}_{1,n}) = \max_{a_2 \in \{-1,1\}} \mathbb{E}_n R_2 | A_2 = a_2, A_1 = \text{sign}(\hat{f}_{1,n}), X_2)$, the objective quantity in Step 3 equates to

$$\mathbb{E}_n \left[\mathbb{E}_n (R_1 + Q_2^*(\hat{f}_{1,n})) I(A_1 = f_1(H_1)) | A_2 = \text{sign}(T_2(\hat{f}_{1,n})) \right].$$

Therefore, with the operator T_1 on $\hat{f}_{2,n}$ introduced, where $\hat{f}_{2,n}$ has been replaced by $\hat{f}_{2,n}^{new}$, we define $\hat{f}_{1,n}^{new} = T_1(T_2(\hat{f}_{1,n}))$, and

$$\begin{aligned}\text{sign}(\hat{f}_{1,n}^{new}) =& \text{sign} \Big[\mathbb{E}_n (R_1 + Q_2^*(\hat{f}_{1,n}) | A_2 = \text{sign}(T_2(\hat{f}_{1,n})), X_1, A_1 = 1) \\ & - \mathbb{E}_n (R_1 + Q_2^*(\hat{f}_{1,n}) | A_2 = \text{sign}(T_2(\hat{f}_{1,n})), X_1, A_1 = -1) \Big].\end{aligned}$$

We thus have

$$V(T_1(T_2(\hat{f}_{1,n})), T_2(\hat{f}_{1,n})) = \max_{a_1 \in \{-1,1\}} \mathbb{E}_n (R_1 + Q_2^*(\hat{f}_{1,n}) | A_2 = \text{sign}(T_2(\hat{f}_{1,n})), X_1, A_1 = a_1).$$

It is straightforward to see that each iteration of the algorithm increases the value function, since

$$V(T_1(T_2(\hat{f}_{1,n})), T_2(\hat{f}_{1,n})) \geq V(\hat{f}_{1,n}, T_2(\hat{f}_{1,n})) \geq V(\hat{f}_{1,n}, \hat{f}_{2,n}).$$

In the beginning, only part of the data, where the second stage assignments matched the estimated treatments, are utilized in finding the first stage strategy. We then continue to the next step and reestimate the second stage strategy, based on the subset where subjects indeed received the estimated first stage assignment. If the optimal dynamic treatment regime has been identified from the beginning, most likely we obtain the same results and the iterations stop. On the other hand, we may see a different second stage strategy produced, resulting in a different subset that can be applied for selecting the first stage therapy. Subsequently, we are able to use the information that was not applied initially for the first stage estimation. As the iterations progress, the results are gradually refined with more of the data until a firm solution has been determined.

4.3 Theoretical Results

In this section, we present the theoretical results for the methods described in Section 4.2, which provide justifications for using backwards outcome weighted learning to find the optimal dynamic treatment regimes.

4.3.1 Fisher Consistency

In the following proposition, we show that by replacing the zero-one loss with the hinge loss in the target function (4.2.1) and solving the resulting optimization problem with the surrogate loss backwards, we obtain a sequence of decision rules that is equivalent to the optimal dynamic treatment regime.

Proposition 4.3.1. *If we minimize (4.2.2) backwards through time* $t = T, T-1, \ldots, 1$ *and obtain a sequence of decision functions* $\{\tilde{f}_T(h_T), \ldots, \tilde{f}_1(h_1)\}$ *by replacing* $V^*_{t+1}(H_{t+1})$ *with* $V_{t+1}(H_{t+1})|_{d=sign(\tilde{f})}$ *at stage t, then* $d^* = sign\{\tilde{f}_1(h_1), \ldots, \tilde{f}_T(h_T)\}$.

Proof. To show this, first note that $V^*_{T+1}(H_{T+1}) = 0$. According to the previous results on Fisher consistency for the single stage problem, we obtain that $d^*_T(h_T) = sign(\tilde{f}_T(h_T))$, and the resulting value function for stage T is indeed $V^*_T(h_T)$. Repeating the arguments for stages $T-1, \ldots, 1$, we have the desired conclusion.

This theorem validates the usage of the hinge loss in the implementation, indicates that the BOWL procedure aims at the optimal dynamic treatment regime directly.

4.3.2 Consistency

The theorem below shows that, as sample size increases, the value of the deduced treatment regimes starting at arbitrary stage t via BOWL converges to the best possible value. To this end, we first define for $t = 1, \ldots, T$,

$$\hat{\mathcal{R}}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) = \mathbb{E}_n \left[\frac{(R_t + \hat{V}_{t+1,n}(\hat{f}_{t+1,n}, \ldots, \hat{f}_{T,n}))I(A_t \neq \text{sign}(\hat{f}_{t,n}(h_t)))}{A_t\pi_t + (1 - A_t)/2} \right| \\ H_t = h_t, A_l = \text{sign}(\hat{f}_{l,n}(H_l)), l = t+1, \ldots, T \Bigg]. \qquad (4.3.1)$$

$$\tilde{\mathcal{R}}_t(h_t, \hat{f}_{t+1,n}, \ldots, \hat{f}_{T,n}) = \min_{f_t} E \left[\frac{(R_t + \hat{V}_{t+1,n}(\hat{f}_{t+1,n}, \ldots, \hat{f}_{T,n}))I(A_t \neq \text{sign}(f_t(h_t)))}{A_t\pi_t + (1 - A_t)/2} \right| \\ H_t = h_t, A_l = \text{sign}(\hat{f}_{l,n}(H_l)), l = t+1, \ldots, T \Bigg]. \qquad (4.3.2)$$

$$\mathcal{R}_t^*(h_t) = \min_{f_t} E \left[\frac{(R_t + V_{t+1}^*(H_{t+1}))I(A_t \neq \text{sign}(f_t(h_t)))}{A_t\pi_t + (1 - A_t)/2} \right| H_t = h_t, \\ A_l = \text{sign}(\hat{f}_{l,n}(H_l)), l = t+1, \ldots, T \Bigg]. \qquad (4.3.3)$$

Note that (4.3.3) follows since R_t and $V_{t+1}^*(H_{t+1})$ are independent of $A_l, l = t+1, \ldots, T$, given H_t. We have the following theorem showing that the value calculated from $(\hat{f}_{1,n}, \ldots, \hat{f}_{T,n})$ converges to the optimal value function. The proof of the theorem can be found in the Appendix 2.

Theorem 4.3.2. *Assume that at stage t, $t = 1, \ldots, T$, we choose a sequence $\lambda_{t,n}$ such that $\lambda_{t,n} \to 0, n_t\lambda_{t,n} \to \infty$. In addition, if the minimizer of $\tilde{\mathcal{R}}_t(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n})$ belongs to the closure of $\limsup_n \mathcal{H}_{k_t}$, then for all distributions P, we have that in probability,*

$$\lim_{n_t \to \infty} \hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) = V_t^*(h_t).$$

The penalization parameter $\lambda_{t,n}, t = 1, \ldots, n$ changes with n. How to select the sequence of $\lambda_{t,n}$ is an important problem. In our simulation study, we apply the common approach of cross validation to choose $\lambda_{t,n}$ at each stage. With appropriately selected tuning parameters, the estimated sequence of treatment regimes approaches the optimal DTR asymptotically.

4.3.3 Risk Bound and Convergence Rate

In addition to asymptotic consistency results, we are interested in how fast the convergence to the optimal value happens. It is also important that the algorithm guarantees a small error when comparing with the optimal. We now derive the convergence rate of $\hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) - V_t^*(h_t)$. Some regularity conditions are required on the data distribution. Specifically, we have a "geometric noise" assumption as follows. Let

$$\eta_t(h_t) = \frac{E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = 1) - E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = -1)}{E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = 1) + E(R_t + V_{t+1}^*(H_{t+1})|H_t = h_t, A_t = -1)} + 1/2. \tag{4.3.4}$$

Then $2\eta_t(h_t) - 1$ is the decision boundary for the optimal ITR in the t^{th} stage. For each stage t, we further define $\mathcal{H}_t^+ = \{h_t \in \mathcal{H}_t : 2\eta(h_t) - 1 > 0\}$, and $\mathcal{H}_t^- = \{h_t \in \mathcal{H}_t : 2\eta(h_t) - 1 < 0\}$. A distance function to the boundary between $\mathcal{H}_t^+$ and $\mathcal{H}_t^-$ is $\Delta(h_t) = \tilde{d}(h_t, \mathcal{H}_t^+)$ if $h \in \mathcal{H}_t^-$, $\Delta(h_t) = \tilde{d}(h_t, \mathcal{H}_t^-)$ if $h_t \in \mathcal{H}_t^+$ and $\Delta(h_t) = 0$ otherwise, where $\tilde{d}(h_t, \mathcal{O})$ denotes the distance of h_t to a set $\mathcal{O}$ with respect to the Euclidean norm. Then the distribution P is said to have geometric noise exponent $0 < q_t < \infty$ (Steinwart and Scovel, 2007), if there exists a constant $C > 0$ such that

$$E\left[\exp\left(-\frac{\Delta(H_t)^2}{\vartheta}\right)|2\eta_t(H_t) - 1|\right] \leq Ct^{q_t p_t/2}, \vartheta > 0. \tag{4.3.5}$$

Additionally, to calculate the convergence rate of the BOWL estimator, we consider the RKHS $\mathcal{H}_{k_t}$, in which the stage t decision function f_t resides, as the space associated with Gaussian Radial Basis Function (RBF) kernels $k_t(h_t, h_t') = \exp(-\sigma_{t,n}^2 \|h_t - h_t'\|^2), h_t, h_t' \in \mathcal{H}_t$. $\sigma_{t,n} > 0$ is a parameter varying with n controlling the bandwidth of the kernel. The Gaussian RBF kernel, a nonlinear kernel, has shown good general performance and has been widely used in various application areas. Because of its flexibility, RBF may summarize the targeted functions better. Moreover, by using the Gaussian RBF kernel, the complexity of $\mathcal{H}_{k_t}$ can be controlled via the empirical L_2-norm, defined as

$$\|f - g\|_{L_2(P_n)} = \left(\frac{1}{n_t} \sum_{i=1}^{n_t} |f(H_{t,i}) - g(H_{t,i})|^2 \right)^{1/2}.$$

For any $\epsilon > 0$, the covering number of a functional class $\mathcal{F}$ with respect to $L_2(P_n)$, $N(\mathcal{F}, \epsilon, L_2(P_n))$, is the smallest number of $L_2(P_n)$ ϵ-balls needed to cover $\mathcal{F}$, where an $L_2(P_n)$ ϵ-ball around a function $g \in \mathcal{F}$ is the set $\{f \in \mathcal{F} : \|f - g\|_{L_2(P_n)} < \epsilon\}$. We have that at stage t, for any $\epsilon > 0$,

$$\sup_{P_n} \log N(B_{\mathcal{H}_{k_t}}, \epsilon, L_2(P_n)) \leq c_{\nu,\delta,p_t} \sigma_{t,n}^{(1-\nu/2)(1+\delta)p_t} \epsilon^{-\nu},$$

where $B_{\mathcal{H}_{k_t}}$ is the closed unit ball of $\mathcal{H}_{k_t}$, and ν and δ are any numbers satisfying $0 < \nu \leq 2$ and $\delta > 0$.

Theorem 4.3.3. *Let the distribution of $(H_t, A_t, R_t), t = 1, \ldots, T$ satisfy condition (4.3.5) with noise exponent $q_t > 0$. Then for any $\delta > 0, 0 < \nu < 2$, there exists a constant C_t (depending on ν, δ, p_t and π_t), such that for all $\tau \geq 1$ and $\sigma_{t,n} = \lambda_{t,n}^{-1/(q_t+1)p_t}$,*

$$Pr^* \left(\hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) \geq V_t^*(h_t) - \sum_{l=t}^{T} 2^{l-t} \epsilon_l \right) \geq 1 - \sum_{l=t}^{T} 2^{l-t} e^{-\tau},$$

where

$$\epsilon_l = C_l \left[\left(\frac{1}{\lambda_{l,n}} \right)^{\frac{2}{2+\nu} + \frac{(2-\nu)(1+\delta)}{(2+\nu)(1+q_l)}} \left(\frac{1}{n_l} \right)^{\frac{2}{2+\nu}} + \left(\frac{1}{\lambda_{l,n}} \right)^{\frac{q_l}{q_l+1}} \frac{\tau}{n_l} + \lambda_{l,n}^{\frac{q_l}{q_l+1}} \right], \tag{4.3.6}$$

and $n_l = \sum_{i=1}^{n} I(A_{l+1} = \hat{f}_{l+1,n}, \ldots, A_T = \hat{f}_{T,n})$. *Specifically,* $n_l = n/2^{T-l}$ *if* $\pi_1 = \ldots = \pi_T = 0.5$.

Theorem 4.3.3 measures the probability that the difference between the value of the estimated DTR and the optimal value is sufficiently small. Furthermore, we can derive the rate of convergence of the estimated values approaching the corresponding targeted optimal values. Each $\epsilon_l, l = 1, \ldots, T$ consists of the estimation error, the first two terms, and the approximation error, the last term. Estimation error reflects the variability from using finite sample sizes, while the candidate function spaces are fixed. Approximation error essentially represents the bias by comparing the best possible result in the selected function spaces with that across all possible spaces. In particular, we can let $q_1 = \ldots = q_T = q$, and choose $\lambda_{t,n}$ for stage t as

$$\lambda_{t,n} = n_t^{-\frac{2(1+q)}{(4+\nu)q+2+(2-\nu)(1+\delta)}},$$

which balances bias and variance. Then the optimal rate for the value of the estimated regimes starting at arbitrary stage t using BOWL is

$$\hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) \geq V_t^*(h_t) - O_p\left(n_t^{-\frac{2q}{(4+\nu)q+2+(2-\nu)(1+\delta)}} \right).$$

In the above formula, δ is a free parameter which can be set arbitrarily close to zero. The geometric noise component q is related to the noise condition regarding the separation between two optimal treatment groups. ν measures the order of complexity for the associated reproducing kernel hilbert space. If q is sufficiently large, which is possible when two optimal treatment groups are well separated, and ν is close to zero, the convergence rate is approximately $n_t^{-1/2}$. Here n_t is the sample size for the stage t treatment estimation.

4.3.4 Improved Rate with Data Completely Separated

Analogous to the single stage setup, we show that a faster convergence rate can be obtained if the data are completely separated. Intuitively, this means patients are more sensitive to different treatments across all stages. Assume that for each stage t,

(A1) $\forall h_t \in \mathcal{H}_t$, $|\eta_t(h_t) - 1/2| \geq \eta_0 > 0$, where $\eta_t(h_t)$ is defined in (4.3.4), and η_t is continuous.

(A2) $\forall h_t \in \mathcal{H}_t$, $\min(\eta_t(h_t), 1 - \eta_t(h_t)) \geq \eta_1 > 0$.

The following theorem gives a faster convergence rate:

Theorem 4.3.4. *Assume that (A1) and (A2) are satisfied. For any* $\nu \in (0, 1)$ *and* $q_t \in (0, \infty), t = 1, \ldots, T$, *let* $\lambda_{t,n} = O(n_t^{-1/(\nu+1)})$ *and* $\sigma_{t,n} = \lambda_{t,n}^{-1/(q_t+1)p_t}$. *Then*

$$V_t^*(h_t) - \hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) = \sum_{l=t}^{T} O_p\left(n_l^{-\frac{1}{\nu}\frac{q_l}{q_l+1}}\right),$$

where $n_l = \sum_{i=1}^{n} I(A_{l+1} = \hat{f}_{l+1,n}, \ldots, A_T = \hat{f}_{T,n})$. *Specifically,* $n_l = n/2^{T-l}$ *if* $\pi_1 = \ldots = \pi_T = 0.5$.

Assuming that the geometric noise components are the same in all the stages, i.e., $q_1 = \ldots = q_T = q$, we obtain for any $t = 1, \ldots, T$

$$V_t^*(h_t) - \hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) = O_p\left(n_l^{-\frac{1}{\nu}\frac{q}{q+1}}\right).$$

Letting q go to ∞ and ν go to zero, the convergence rate for $V_t^*(h_t) - \hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n})$ at stage t is almost n_t^{-1}.

4.4 Simulation Studies

To assess the performance of the proposed methods, we conduct simulations under various scenarios. We consider a two stage clinical trial. 50 dimensional baseline covariates $X_{1,1}, \ldots, X_{1,50}$ are generated according to $N(0, 1)$. Treatments A_1, A_2 were randomly generated from $\{-1, 1\}$ with equal probability 0.5. The resulting outcomes for different stages, R_1 and R_2, vary under different settings stated below. Thus we observe data of the form $(X_1, A_1, R_1, X_2, A_2, R_2)$ on each patient, where X_2 denotes the variables observed prior to stage 2. Specifically, histories available at each stage are: $H_1 = X_1$, and $H_2 = (X_1, A_1, R_1, X_2)$. In each scenario, we simulate a validation data set of sample size 10000, where the expected summation of outcomes $E(R_1 + R_2)$ are evaluated. 500 replications of training data sets are also simulated, with sample sizes varying from 100 to 200, 400 and 800. For illustration, we present three scenarios,

1. Stage 1 outcome $R_1 = 0$, and stage 2 outcome R_2 is generated according to $N(-0.5A_1 + 0.5A_2 + 0.5A_1A_2, 1)$.

2. Stage 1 outcome R_1 is generated according to $N(0.446X_{1,3}A_1, 1)$, and stage 2 outcome R_2 is generated according to $N(((X_{1,1}^2 + X_{1,2}^2 - 0.2)(0.5 - X_{1,1}^2 - X_{1,2}^2) + R_1)A_2, 1)$.

3. A more complex model with intermediate variables after stage 1, specifically,

   (a) Stage 1 outcome R_1 is generated according to $N((1 + 1.5X_{1,3})A_1, 1)$.

   (b) Two intermediate variables are generated with $X_{2,1} \sim N(1.25X_{1,1}A_1, 1)$, and $X_{2,2} \sim N(-1.75X_{1,2}A_1, 1)$.

(c) Stage 2 outcome R_2 is generated according to $N((0.5+R_1+0.5A_1+0.5X_{2,1}-0.5X_{2,2})A_2, 1)$.

Scenario 1 is a toy example with simple models studied in Chakraborty et al. (2009). Here we add 50 dimensional baseline covariates into the setup, which are pure noise and have no effects on the outcomes. There are no time-varying covariates involved in scenario 2, and a non-linear relationship exists between baseline covariates and the optimal stage 2 treatment. Also, intermediate outcomes R_1 play a role in determining the second stage outcomes. We incorporate two time-varying covariates in scenario 3, i.e., the values of the first and second baseline variables will change after stage 1.

Different methods are compared, including Q-learning with linear regression, backwards outcome weighted learning (BOWL) with linear kernel and iterative outcome weighted learning (IOWL). The analysis model for Q-learning is $Q_j(H_j, A_j) = \beta_j H_j + (\psi_j H_j)A_j, j = 1, 2$. Considering that the outcomes are modeled with linear regression in Q-learning, we carry out the proposed outcome weighted learning methods utilizing linear kernels for illustration, and do not further explore the use of Gaussian kernels hereafter. We follow the BOWL procedures described in Section 4.2.2, where the optimal stage 2 treatments are obtained via a weighted support vector machine technique based on history H_2, with the optimization target defined in (4.2.2). The estimation of the optimal treatment in stage 1 is then carried out using the history information H_1 on the subset of patients whose assignments A_2 are consistent with the estimated decisions $\hat{d}_2$. We use 5-fold cross validation to select tuning parameters in each stage. The data is partitioned into 5 subsets. Each time 4 subsets are used as the training data for treatment estimation using OWL, while the remaining set is used as the validation data for calculating the value of the estimated rule. The process is repeated 5 times and we average the value obtained each time. In particular, using the linear kernel in the implementation of the weighted support vector machine, we choose the penalization parameter $\lambda_{t,n}$ in (4.2.2) for stage $t, t = 1, 2$, which maximizes the average of 5 estimated values, among a pre-defined set of values in each stage. The IOWL updates the estimated decisions back and forth between two stages. As mentioned in Section 4.2.2, in each iteration, the one stage rule is rediscovered using the weighted support machine technique, based on the group of patients receiving the recommended treatment for the other stage. Here again, cross validation is utilized to select the required tuning parameter via a grid search. The iterative procedure stops upon stabilization of the value functions or reaching the maximum number of iterations, set at 20 in our simulations.

We can compute the exact values of the value function when estimated dynamic treatment regimes are applied to the large simulated validation set using different methods. Subsequently, we plot the probability distribution of the obtained values in Figure 4.1, 4.3 and 4.5, which are estimated based on a Gaussian kernel function with the bandwidth set to 0.5. We also plot quantiles of the differences between estimated and optimal values against quantiles of the estimated values from Q-learning linear, see Figure 4.2, 4.4 and 4.6 for Scenario 1, 2 and 3, respectively. For the first scenario with a simple effects model, Q-learning performs worse comparatively, especially with smaller sample sizes. There is little difference between different outcome weighted learning based methods as can be seen by observing that the kernel density estimates almost overlap in Figure 4.1. Most of the time, they correctly identify the optimal decision rules and reach the exact best values, see Figure 4.2. For Scenario 2, the linear relationship between covariates and desired treatments is not valid with a fairly non-linear effect existing in the second stage. In this situation, Q-learning linear may never achieve the correct decision boundary. As shown in Figure 4.3, Q-learning linear tends to estimate the wrong target since the mean of the distribution deviates from the optimal value substantially. This is anticipated because the linear model for the response can never catch the truth of non-linear treatment effects. Outcome weighted learning methods outperform Q-learning linear, and the gain becomes more pronounced with increasing sample sizes. The strength of iterative outcome weighted learning is demonstrated in this example. By taking advantage of this iterative process cycling over the complete data set, it improves the decision from BOWL with better precision. Behaviors of the two outcome based learning methods are consistent in the sense that different approaches most of the time lead to the same value, which is close to the truth: see the partially overlapped lines in Figure 4.4. Scenario 3 takes evolving variables into consideration. Still, Q-learning linear gives worse performances, especially with small sample sizes.

Tables 4.1, 4.2, and 4.3 show the mean values of the estimated DTR on the testing set over 500 runs. Among all scenarios, Q-learning does not achieve good results until sample sizes reach 400, sometimes even 800. Implementing outcome weighted learning in an iterative fashion render the results more stabilized. Table 4.4 reports the percentages when the estimated values using BOWL or IOWL are higher than or equal to those using Q-learning. It turns out that most of the times, outcome weighted learning based methods yield better dynamic treatment regimes on average, even if occasionally mean results can be strongly affected by the outliers with low values.

4.5 Data Analysis: Smoking Cessation Study

The smoking cessation study consists of two stages. The purpose of stage 1 of this study (Project Quit), lasting for 6 months, was to find an optimal multicomponent behavioral intervention to help adult smokers quit smoking; and among the participants of Project Quit, the following 6-month stage 2 (Forever Free) was conducted to help those who already quit stay quit, and help those who failed continue the quitting process. However, many participants from Project Quit did not continue to Forever Free, where only 479 out of 1848 subjects decided to continue. For more details, see Chakraborty et al. (2009).

The baseline covariates considered in stage 1 include 10 variables, described in Table 4.5. We consider Story as the only stage one treatment variable A_1, coded 1 and -1, representing two levels with high vs. low tailoring depth, i.e., whether or not the story is tailored to the individual. At stage 2, FFArm denotes the treatment effects A_2, where 1 indicating subjects assigned into the treatment group and -1 indicating subjects assigned into the control group. Note that there were originally 4 different treatment groups. However, they were combined for the analysis since there were little differences between them.

Total number of months without smoking is taken as the outcome of interest in the study. The stage 1 outcome R_1, nonsmoking months in the Project Quit period, is collected at 6 months from the date of randomization, and the stage 2 outcome R_2, nonsmoking months in the Forever Free period, is obtained at 6 months from the date of stage 2 randomization (i.e., 12 months from the date of stage 1 randomization). Only 281 subjects completed the stage 2 six-month survey, and there were some missingness in the collected variables. By excluding all the missingness, we have 193 observations for stage 2.

We are mainly interested in examining the performance of different outcome weighted learning based approaches, i.e., BOWL and IOWL. Also, Q-learning is conducted as a competitor. For the implementation of Q-learning, we need to posit a model for each decision point. We first incorporate all the history covariates and covariate-treatment interactions into the prediction models for Q-functions, and denote this approach as Q-learning complete. Furthermore, it is found that the effect of story is thought to interact with education, that highly tailored level of story is more effective for participants with lower education. Hence, another strategy is to consider a parsimonious model which only includes the interaction term between education ($X_{1,2}$) and story (A_1) in the first stage. We call it "Q-learning simple". To avoid potential problems from overfitting, we compute value estimates from a cross validated procedure, which is repeated 100 times. Specifically, each time we randomly split the data into 5 roughly equally sized parts. 4 out of 5 parts of the data are used as the training set, on which different methods are applied to construct the optimal dynamic treatment regimes, and the remaining part is retained for validation by calculating values of the obtained estimates. The process is repeated 5 times and the averages of the computed values are recorded.

Results of 100 cross validated values are shown in Figure 4.7. It can be seen that IOWL gives larger cross validated values most of the times, indicating potentially better sequences of treatments for the patients who can benefit from a longer time without smoking. Performances of BOWL and Q-learning simple are close, while Q-learning complete method is the worst comparatively, probably due to the large set of interactions to account for. Indeed, Table 4.6 gives the mean of the cross validated values across 100 times, where IOWL yields the highest value on average.

4.6 Discussion

It is critical to recognize that the presented approaches are developed for the discovery of optimal dynamic treatment regimes. In contrast to the typical randomized clinical trials, which are conducted to confirm the efficacy of the new treatments, the SMART designs mentioned at the beginning are devised for exploratory purposes in developing optimal DTRs. However, a confirmatory trial with a phase III structure can be used for followed up to validate the superiority of the optimal adaptive treatment strategies compared to existing therapies.

We have proposed novel methodologies for identifying the optimal dynamic treatment regimes from outcome weighted learning perspectives. The presented methods, formulated in a nonparametric framework, are computational efficient, easy and intuitive to apply, and can effectively handle the potential complex relationship between sequential treatments and prognostic or intermediate variables in the multi-decision problem. For backwards outcome weighted learning, we conduct the estimation procedure backwards through time, i.e., from the last time point back to the beginning of the study, without knowing the underlying distribution of the entire process as is required in dynamic programming. Instead of modeling the Q-function at each stage, we directly maximize the value function stepwise. An iterative procedure can be further introduced which potentially yield more stabilized and accurate results. The convergence rates derived for different OWL based methods, are essentially the same under prespecified conditions on separation, smoothness and complexity of the approximation spaces. Clearly, the proposed methodologies, augmenting the current literature in dynamic treatment regimes from a different point of view, can be powerful and promising tools for improving long term health outcomes when managing chronic diseases.

Table 4.1: Mean Values of the Estimated DTR for Scenario 1, where the Optimal Value Equals 0.500

Q-learning	BOWL	IOWL
0.223	0.496	0.488
0.154	0.499	0.494
0.487	0.499	0.498
0.500	0.500	0.500

Table 4.2: Mean Values of the Estimated DTR for Scenario 2, where the Optimal Value Equals 7.095

Q-learning	BOWL	IOWL
0.639	4.933	5.413
0.679	4.872	5.899
3.992	4.253	5.800
5.548	4.410	6.520

Table 4.3: Mean Values of the Estimated DTR for Scenario 3, where the Optimal Value Equals 3.750

Q-learning	BOWL	IOWL
0.839	2.633	2.466
0.575	2.837	2.642
2.347	3.052	2.739
3.016	3.198	2.955

Table 4.4: Proportions of Higher or Same Values using BOWL and IOWL Compared to using Q-learning

Scenario 1		Scenario 2		Scenario 3	
BOWL	IOWL	BOWL	IOWL	BOWL	IOWL
100.0%	99.2%	82.0%	87.4%	94.2%	91.6%
100.0%	99.8%	81.0%	90.8%	98.0%	97.0%
96.8%	96.6%	60.8%	85.2%	91.4%	81.0%
96.4%	97.0%	42.4%	94.3%	76.6%	37.4%

Table 4.5: Baseline and Intermediate Variables for Smoking Cessation Study

Baseline Variables	
$X_{1,1}$	Age
$X_{1,2}$	Education ($\leq$ high school vs. $>$ high school)
$X_{1,3}$	Gender (Male vs. Female)
$X_{1,4}$	Race1 (Black vs. Non Black)
$X_{1,5}$	Race2 (White vs. Non White)
$X_{1,6}$	Baseline motivation to quit smoking*
$X_{1,7}$	Baseline self efficacy**
$X_{1,8}$	Average number of cigarettes smoked per day at baseline
Intermediate Variables	
$X_{2,1}$	Motivation to quit smoking at 6 months*
$X_{2,2}$	Self efficacy at 6 months**

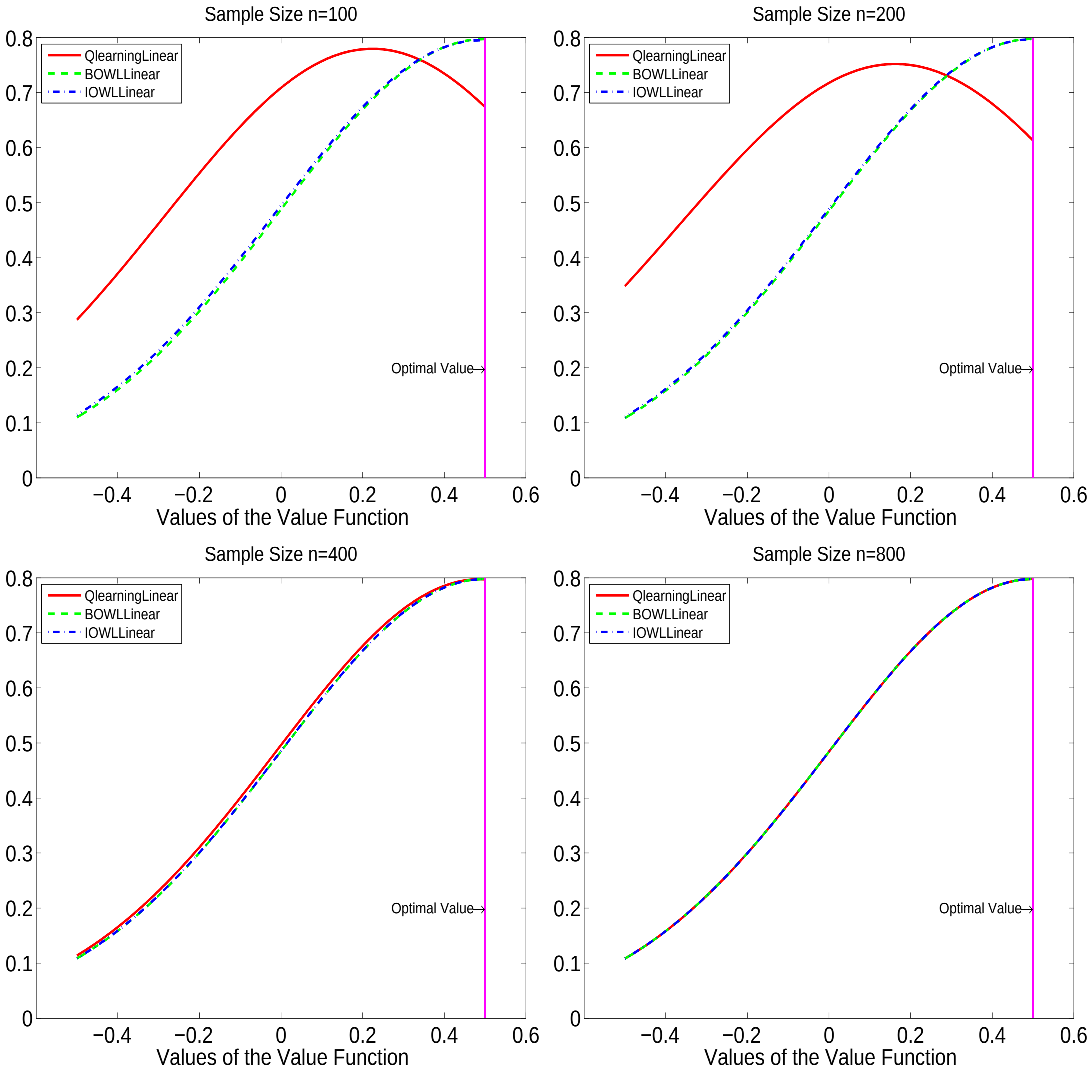
*: originally scaled from 1 to 10 and categorized subjects with 1-5 into 0 and 6-10 into 1.

**: originally scaled from 1 to 10 and categorized subjects with 1-5 into 0 and 6-10 into 1.

Table 4.6: Mean Cross Validated Values using Different Methods

BOWL	IOWL	Q-learning Complete	Q-learning Simple
1.726	1.886	1.259	1.695

Fig. 4.1: Kernel Density Approximations of Estimated Values for Scenario 1



The vertical magenta line represents the optimal value with $V^* = 0.5$ under Setting 1.

Fig. 4.2: Quantile-Quantile Plot of Estimated Values for Scenario 1

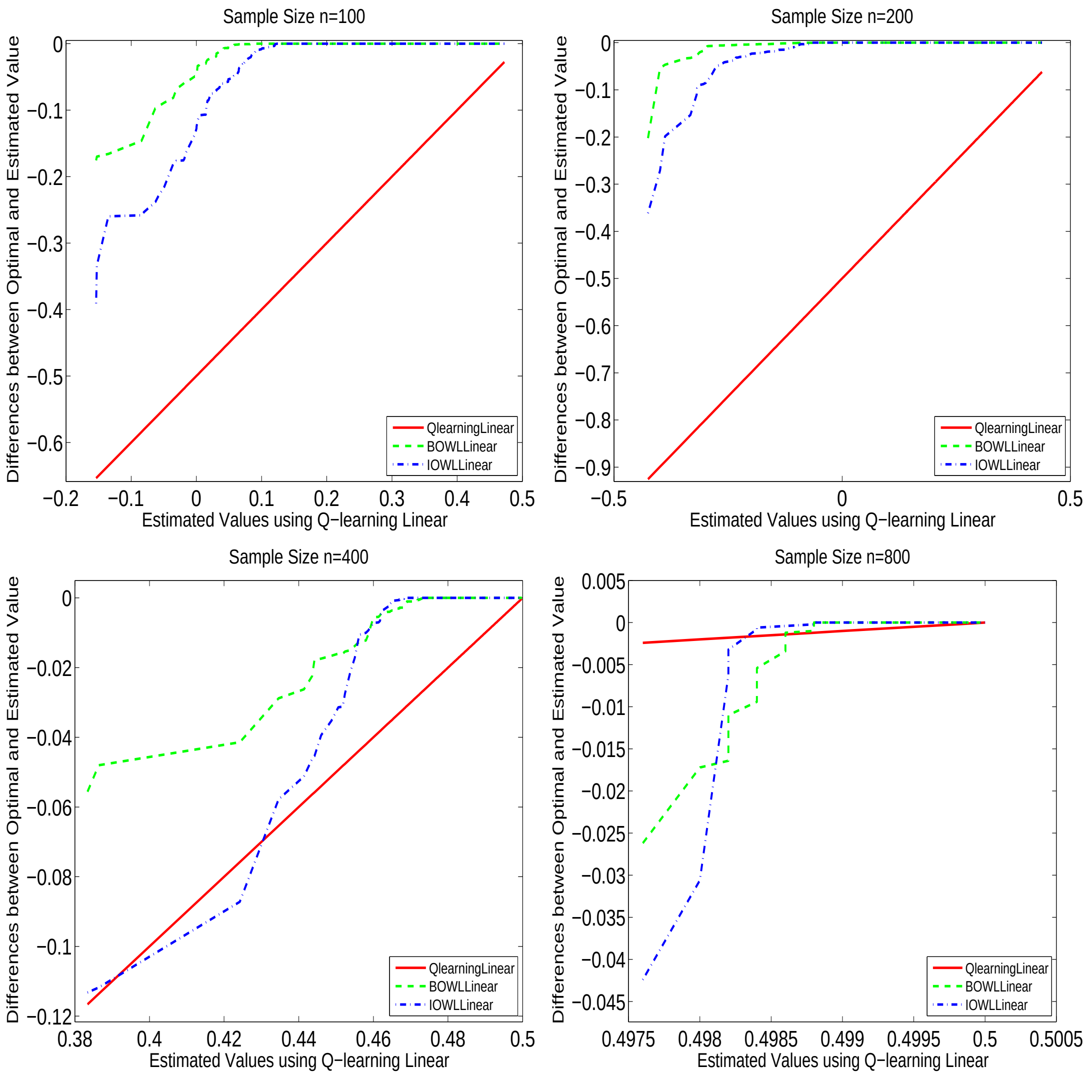
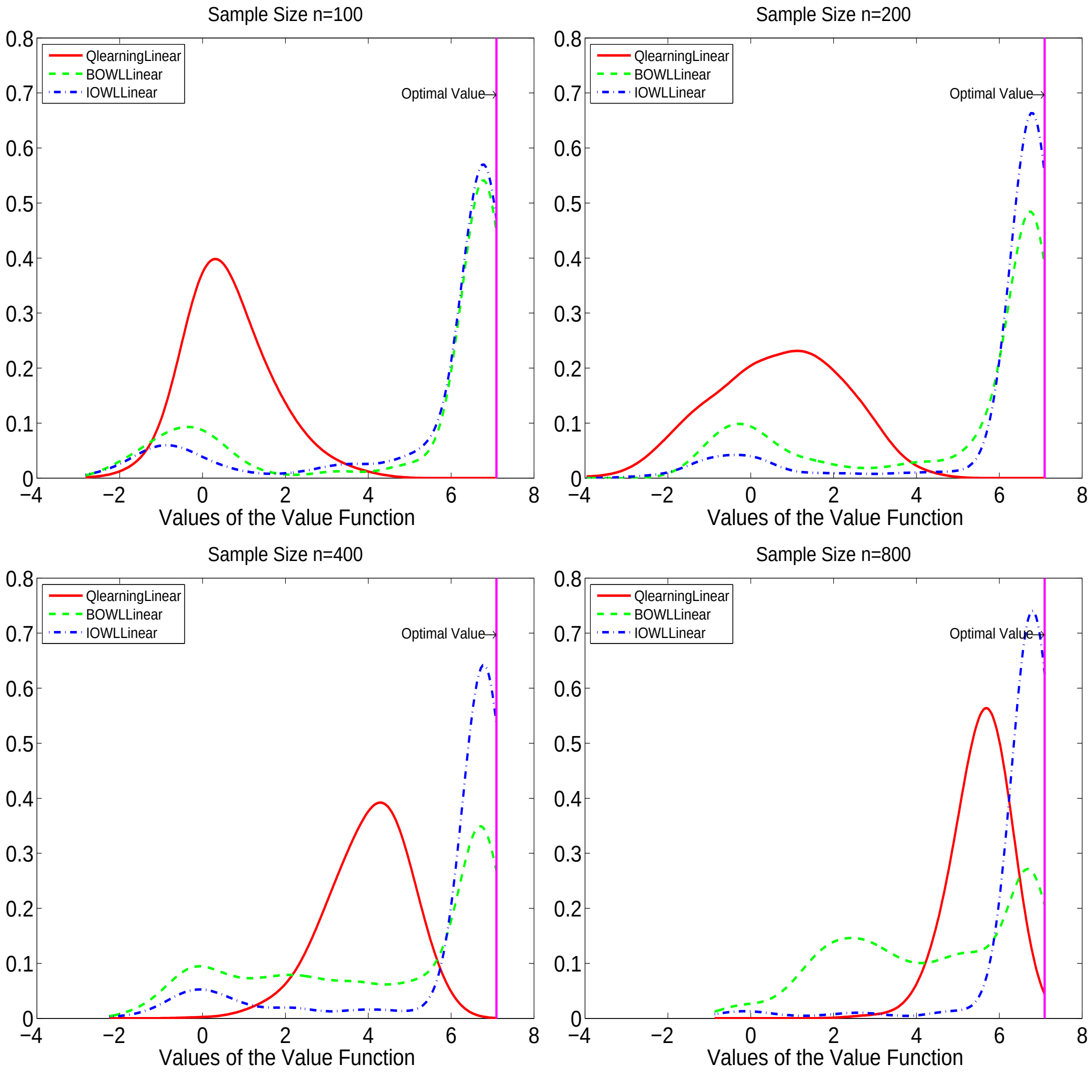


Fig. 4.3: Kernel Density Approximations of Estimated Values for Scenario 2



The vertical magenta line represents the optimal value with $V^* = 7.095$ under Setting 2.

Fig. 4.4: Quantile-Quantile Plot of Estimated Values for Scenario 2

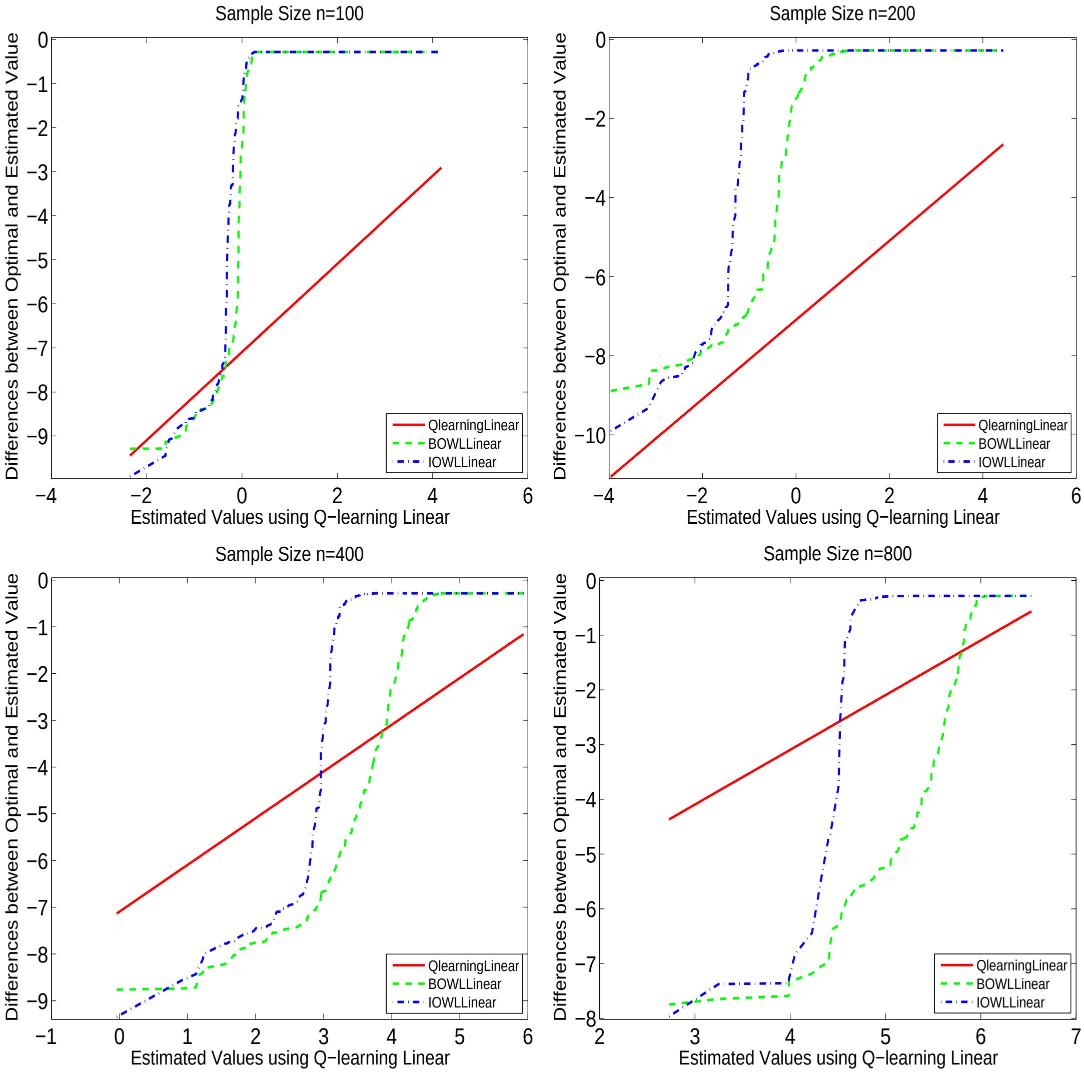
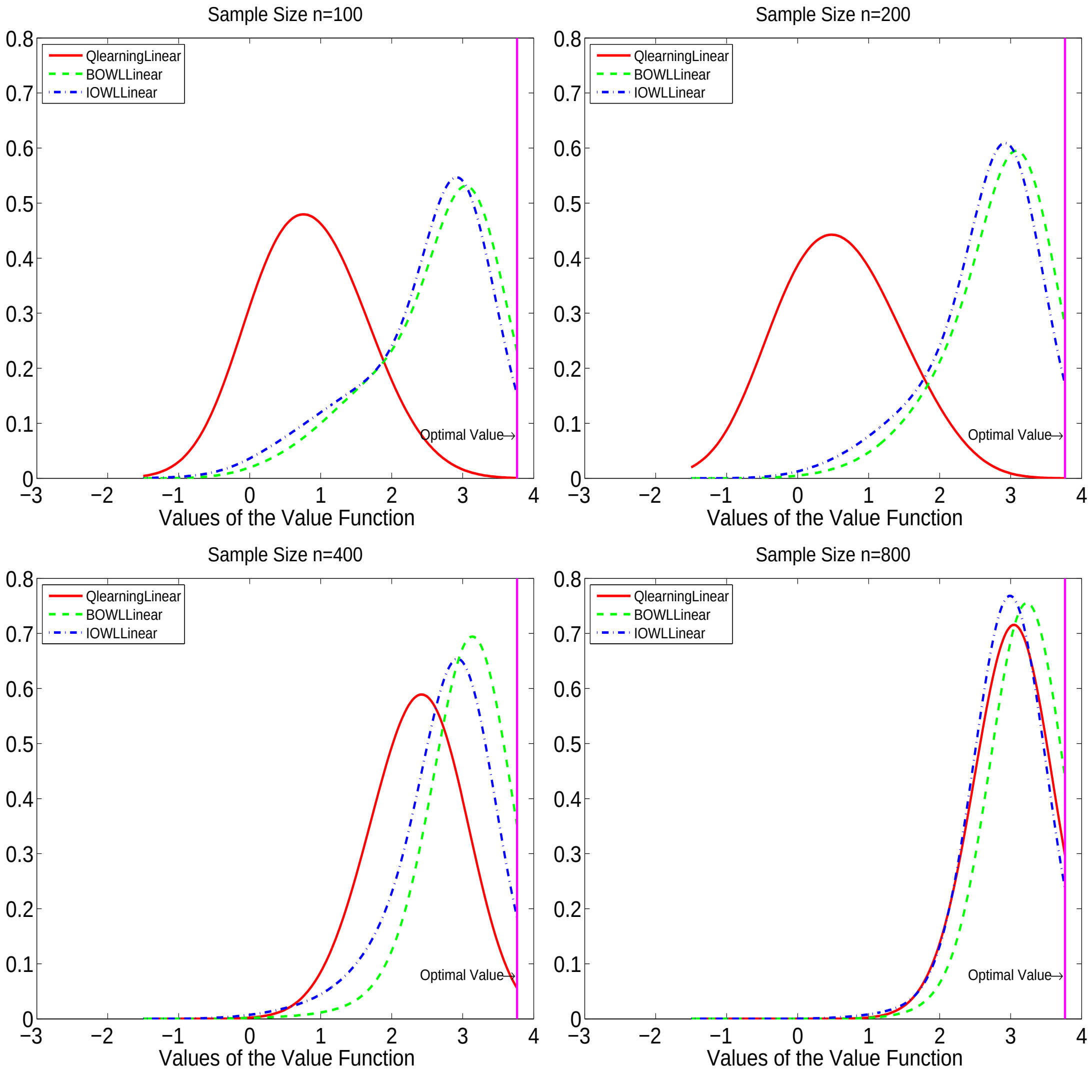


Fig. 4.5: Kernel Density Approximations of Estimated Values for Scenario 3



The vertical magenta line represents the optimal value with $V^* = 3.766$ under Setting 3.

Fig. 4.6: Quantile-Quantile Plot of Estimated Values for Scenario 3

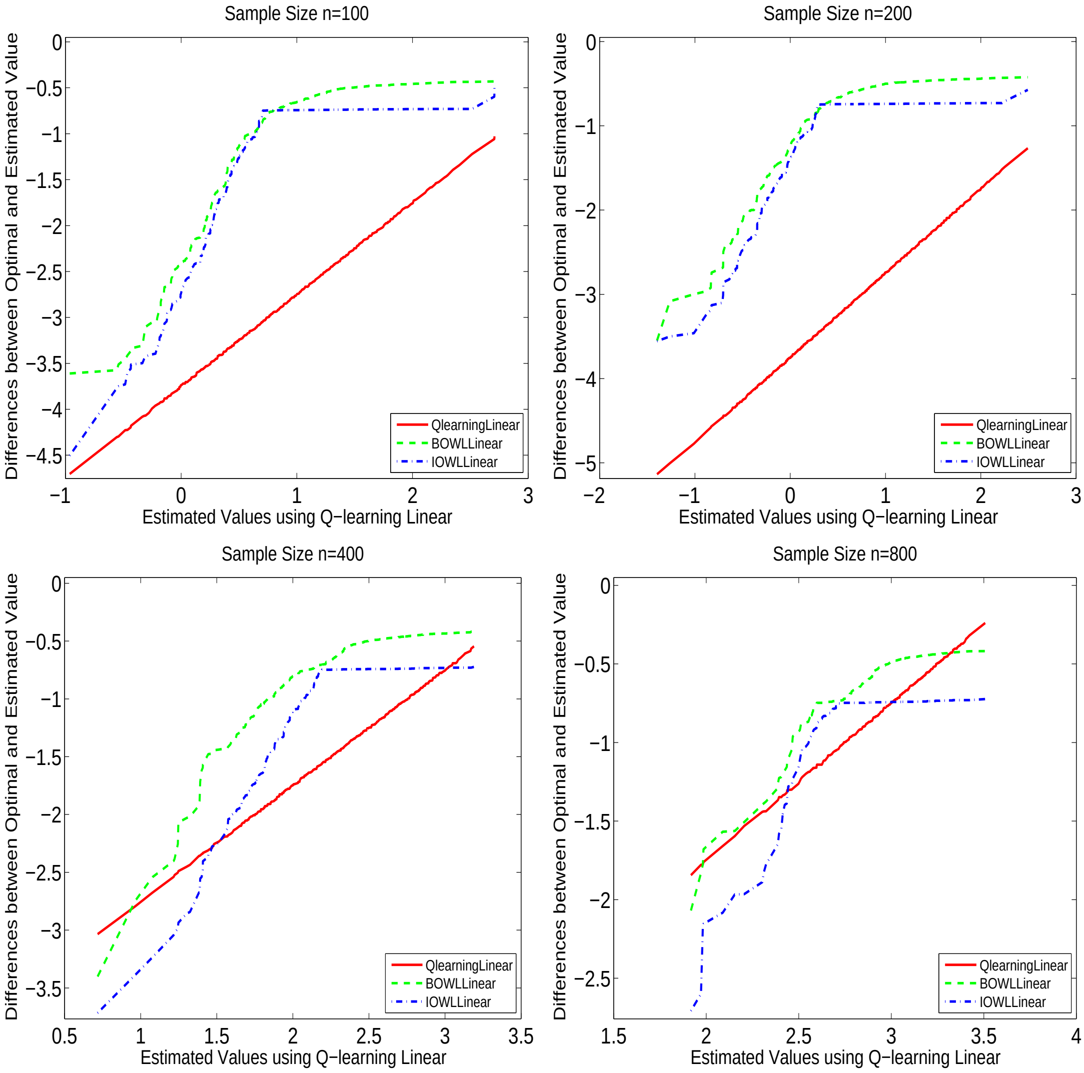
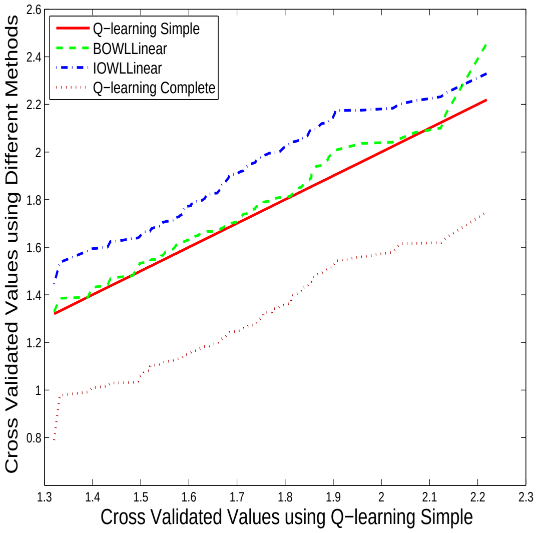


Fig. 4.7: Quantile-Quantile Plot of Cross Validated Values



Chapter 5

Future Work

This dissertation investigates the application of statistical learning in public health surveillance and personalized medicine. In this chapter, we briefly discuss several problems which are worth pursuing in the future.

5.1 Identifying Potential Causes in Public Health Surveillance

It is important not only to detect potential disease outbreaks but also to evaluate them. When multiple outbreaks are observed, we can next investigate the risk factors associated with such outbreaks. As we mentioned, some demographic characteristics of patients come with surveillance data. Other risk factors of interest include sociodemographic variables, geographic information, and environmental influences. Using normal periods as the control, we can assess the significance of each potential risk factor. If change in the surveillance processes is due to some external covariate processes, such as weather, pollution, etc., we expect to detect changes in the feature space considering the potential association. Assuming that some features or attributes may influence the surveillance processes, we also want to find the associated anomalies in the feature space, while we are proposing methodology for detecting space-time anomalies. By observing the historical data, we attempt to discover the relevant covariate processes.

To find associated aberrations of weather/pollution data for different alarmed days at different locations, one approach for mining exploration is to conduct a review of available records to see if one or more possible causes emerge, such as temperature, levels of ozone (ppb), levels of $\mathrm{PM}_{2.5}(ug/m^3)$ and levels of $\mathrm{PM}_{10}(ug/m^3)$, etc. Similarly, we can use local linear methods to obtain estimation for the regular pattern of measurements of interest. Let y denote the variable of interest. For location s on day t, we solve for $\hat{\beta}(s,t)$, where

$$\hat{\beta}(s,t) = \arg\min \sum_{i,j} K_{i,j}(s,t)(y(s_i,t_j) - \beta_0(s,t) - \beta_1(s,t)(t_j - t))^2.$$

Once we obtain the expected pattern from historical data, we conduct residual analysis for screening, including detrended residuals and differenced detrended residuals. Since the measurements are continuous, we define detrended residuals as $y(s,t) - \hat{y}(s,t)$, while differenced detrended residuals are derived via time series modeling. How to choose the order of the time series model depends specifically on data analyzed using the AIC criteria. This aberration analysis procedure enables us be informed of outliers in explanatory variables. We can match these findings with identified alarms from surveillance.

5.2 Extensions in Optimal DTRs Discovery

One important generalization is methodology development for right-censored survival data. Within the Q-learning framework, Zhao et al. (2011a) applied support vector regression to fit the Q-function, which can accommodate right censoring. Goldberg and Kosorok (2012) developed methodology for the multi-decision problem where survival times are outcomes of interests with censoring. They allowed flexible number of stages for different patients depending on disease progression and failure event time. Finite sample bounds are obtained on the generalization error of the estimated DTRs using the proposed Q-learning algorithm. Outcome weighted learning approaches, however, have not been adapted to censored data setting. It is of great interests to pursue this area.

With the advances in new high-throughput technologies, we have encountered challenges arising from huge bodies of data with increasing complexity. Valid methods should be developed to tackle the problem for ultra-high dimensional predictor spaces. An optimal dynamic treatment regime discovery platform that mines variable selection techniques will be more useful, in terms of the simpler and thus more interpretable decision rules. Recall that the proposed OWL is based on a weighted SVM which minimizes the weighted hinge loss function subject to an l_2 penalty. If the dimension of the covariate space is sufficiently large, not all the variables would be essential for optimal ITR construction. By eliminating the unimportant variables from the rule, we could simplify interpretations and reduce health care costs by only requiring collection of a small number of significant prognostic variables. For standard SVM, the l_1 penalty has been shown to be effective in selecting relevant variables via shrinking small coefficients to zero (Bradley and Mangasarian, 1998; Zhu et al., 2003). It outperforms the l_2 penalty when there are many noisy variables and sparse models are preferred. Other forms of penalty have been proposed such as the F_∞ norm (Zou and Yuan, 2008) and the adaptive l_q penalty (Liu et al., 2007). In the future, we can examine use of these sparse penalties in the OWL method. It is likely that multiple, usually more than two, active treatments are available at each decision points. Sometimes we need to determine the dosage level for different patients. Therefore, extensions to multicategory or continuous treatments should be considered for practical reasons.

Conducting statistical inference for dynamic treatment regimes is meaningful to address scientific questions such as "How much confidence do we have in concluding that the obtained optimal dynamic treatment regime is the best compared to other strategies?" or "How many patients are needed in a SMART in order to guarantee that we will obtain a DTR very close to the optimal one, using the proposed methodology?" Efforts have been made to construct confidence intervals for the parameters in the Q-function, with main challenges coming from nonregularity due to the non-differentiability of the max operator (Robins, 2004; Chakraborty et al., 2009; Laber et al., 2011). We have shown that the proposed OWL based methods lead to a reasonably low bias in estimating the optimal DTR and have derived the finite sample bounds for the difference between the expected cumulative outcome using the estimated DTR and that of the optimal one. We believe that this article paves the way for further developments in finding the limiting distribution of the value function and calculating the required sample sizes for the multi-decision problem.

In this dissertation, we have only considered data generated from SMART designs. However, dynamic treatment regimes can be estimated from observational studies. In this setting, the assumption of 'no unmeasured confounders' may be violated. The proposed methods should be applicable by using techniques such as propensity scores (Rosenbaum and Rubin, 1983). Accordingly, other aforementioned issues could be investigated in the context of observational studies as well.

5.3 Concluding Remarks

Recent developments in statistical learning offer a host of new research opportunities. This dissertation has investigated problems related to the two general areas of public health surveillance and personalized medicine, using state-of-the-art statistical learning methods together with semi- and non- parametric modeling techniques. We believe that the current work opens appealing avenues for additional developments, and we are well positioned to continue addressing the potential questions and concerns in the near future.

Appendix 1: Chapter 3 Proofs

Proof of Theorem 3.3.2

We consider the case where rewards are discrete. Arguments for the continuous rewards setting follow similarly. Let $\eta_r(x) = p(A = 1|R = r, X = x)$ and $q_r(x) = rp(R = r|X = x)$. We can write

$$\begin{aligned}
\mathcal{R}(f) &= E\left[\sum_r rp(R = r|X)E\left(\frac{I(A \neq \text{sign}(f(X)))}{A\pi + (1-A)/2}\Big|R = r, X\right)\right] \\
&= E\left[\sum_r q_r(X)\left(\frac{\eta_r(X)}{\pi}I(\text{sign}(f(X)) \neq 1) + \frac{1-\eta_r(X)}{1-\pi}I(\text{sign}(f(X)) \neq -1)\right)\right] \\
&= E\left[c_0(X)(\eta(X)I(\text{sign}(f(X)) \neq 1) + (1-\eta(X))I(\text{sign}(f(X)) \neq -1))\right], \quad (5.3.1)
\end{aligned}$$

where $c_0(x) = \sum_r q_r(x)[\eta_r(x)/\pi + (1-\eta_r(x))/(1-\pi)]$, and $\eta(x)$, defined previously in (3.3.3), is equal to $\sum_r q_r(x)\eta_r(x)/\pi c_0(x)$. Similarly,

$$\mathcal{R}_\phi(f) = E\left[c_0(X)(\eta(X)\phi(f(X)) + (1-\eta(X))\phi(-f(X)))\right].$$

We define $C(\eta, \alpha) = \eta\phi(\alpha) + (1-\eta)\phi(-\alpha)$. Then the optimal ϕ-risk satisfies

$$\mathcal{R}_\phi^* = E\left[c_0(X)\inf_{\alpha\in\mathbb{R}} C(\eta(X), \alpha)\right]$$

and

$$\mathcal{R}_\phi - \mathcal{R}_\phi^* = E\left[c_0(X)\left(C(\eta(X), f(X)) - \inf_{\alpha\in\mathbb{R}} C(\eta(X), \alpha)\right)\right].$$

By a result in Bartlett et al. (2006) for a convexified transform of hinge loss, we have

$$2\eta - 1 = \inf_{\alpha:\alpha(2\eta-1)\leq 0} C(\eta,\alpha) - \inf_{\alpha\in\mathbb{R}} C(\eta,\alpha). \tag{5.3.2}$$

Thus, according to (5.3.1) and (5.3.2), we have

$$\begin{aligned}
&\mathcal{R}(f) - \mathcal{R}^* \leq E\left(I(\text{sign}(f(X)) \neq \text{sign}[c_0(X)(\eta(X)-1/2)])\,|c_0(X)(2\eta(X)-1)|\right)\\
&= E\Bigg[c_0(X)I(\text{sign}(f(X)) \neq \text{sign}[c_0(X)(\eta(X)-1/2)])\Bigg(\inf_{\alpha:\alpha(2\eta(X)-1)\leq 0} C(\eta(X),\alpha) - \inf_{\alpha\in\mathbb{R}} C(\eta(X),\alpha)\\
&\leq E\left[c_0(X)\left(C(\eta(X),f(X)) - \inf_{\alpha\in\mathbb{R}} C(\eta(X),\alpha)\right)\right]\\
&= \mathcal{R}_\phi(f) - \mathcal{R}_\phi^*.
\end{aligned}$$

The last inequality holds because we always have $C(\eta(x), f(x)) \geq \inf_{\alpha\in\mathbb{R}} C(\eta(x),\alpha)$ on the set where $\text{sign}(f(x)) = \text{sign}[c_0(x)(\eta(x)-1/2)]$ and $C(\eta(x), f(x)) \geq \inf_{\alpha:\alpha(2\eta(x)-1)\leq 0} C(\eta(x),\alpha)$ when $\text{sign}(f(x)) \neq \text{sign}[c_0(x)(\eta(x)-1/2)]$.

Proof of Theorem 3.3.3

Define $L_\phi(f) = R\phi(Af)/(A\pi + (1-A)/2)$. By the definition of $\hat{f}_n$, we have for any $f \in \mathcal{H}_k$,

$$\mathbb{P}_n\left(L_\phi(\hat{f}_n)\right) \leq \mathbb{P}_n\left(L_\phi(\hat{f}_n) + \lambda_n \left\|\hat{f}_n\right\|^2\right) \leq \mathbb{P}_n\left(L_\phi(f) + \lambda_n \|f\|^2\right),$$

where $\mathbb{P}_n$ denotes the empirical measure of the observed data. Thus $\limsup_n \mathbb{P}_n(L_\phi(\hat{f}_n)) \leq \mathbb{P}(L_\phi(f))$. It leads to $\limsup_n \mathbb{P}_n(L_\phi(\hat{f}_n)) \leq \inf_{f\in\bar{\mathcal{H}}_k} \mathbb{P}(L_\phi(f))$. Theorem 3.3.3 holds if we can show $\mathbb{P}_n(L_\phi(\hat{f}_n)) - \mathbb{P}(L_\phi(\hat{f}_n)) \to 0$ in probability.

To this end, we first obtain a bound for $\|\hat{f}_n\|_k^2$. Since $\mathbb{P}_n(L_\phi(\hat{f}_n)) + \lambda_n\|\hat{f}_n\|^2 \leq \mathbb{P}_n(L_\phi(f)) + \lambda_n\|f\|_k^2$ for any $f \in \mathcal{H}_k$, we can select $f = 0$ to obtain

$$\|\hat{f}_n\|_k^2 \leq \frac{1}{\lambda_n}\frac{1}{n}\sum \frac{R_i}{\pi_i}\phi(0) \leq \frac{2}{\lambda_n}\frac{E(R)}{\min\{\pi, 1-\pi\}}.$$

Let $M = 2E(R)/\min\{\pi, 1-\pi\}$ so that the $\mathcal{H}_k$ norm of $\sqrt{\lambda_n}\hat{f}_n(X)$ is bounded by $\sqrt{M}$. Note that the class $\{\sqrt{\lambda_n}f : \|\sqrt{\lambda_n}f\|_k \leq \sqrt{M}\}$ is contained in a Donsker class. Thus, $\left\{\sqrt{\lambda_n}L_\phi(f), \|\sqrt{\lambda_n}f\|_k \leq \sqrt{M}\right\}$ is also P-Donsker because $(1 - Af(X))^+$ is Lipschitz continuous with respect to f. Therefore,

$$\begin{aligned}&\sqrt{n}(\mathbb{P}_n - \mathbb{P})L_\phi(\hat{f}_n)\\ =&\sqrt{\lambda_n^{-1}}\sqrt{n}(\mathbb{P}_n - \mathbb{P})\left[\frac{R}{A\pi + (1-A)/2}\left(\sqrt{\lambda_n} - A\sqrt{\lambda_n}\hat{f}_n(X)\right)^+\right] = O_p\left(\sqrt{\lambda_n^{-1}}\right).\end{aligned}$$

Consequently, from $n\lambda_n \to \infty$, $\mathbb{P}_n(L_\phi(\hat{f}_n)) - \mathbb{P}(L_\phi(\hat{f}_n)) \to 0$ in probability.

Proof of Theorem 3.3.4

First, we have

$$\begin{aligned}&\mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi^* \leq \lambda_n\|\hat{f}_n\|_k^2 + \mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi^*\\ &\leq \left[\lambda_n\|\hat{f}_n\|_k^2 + \mathcal{R}_\phi(\hat{f}_n) - \inf_{f\in\mathcal{H}_k}(\lambda_n\|f\|_k^2 + \mathcal{R}_\phi(f))\right] + \left[\inf_{f\in\mathcal{H}_k}(\lambda_n\|f\|_k^2 + \mathcal{R}_\phi(f) - \mathcal{R}_\phi^*)\right].\end{aligned} \tag{5.3.3}$$

We will bound each term on the right-hand-side separately in the following arguments.

For the second term on the right-hand-side of (5.3.3), we use Theorem 2.7 in Steinwart and Scovel (2007) to conclude that

$$\inf_{f\in\mathcal{H}_k}\left(\lambda_n\|f\|_k^2 + \mathcal{R}_\phi(f) - \mathcal{R}_\phi^*\right) = O\left(\lambda_n^{q/(q+1)}\right), \tag{5.3.4}$$

when we set $\sigma_n = \lambda_n^{-1/(q+1)d}$.

Now we proceed to obtain a bound for the first term on the right-hand-side of (5.3.3). To do this, we need the useful Theorem 5.6 of Steinwart and Scovel (2007) presented below:

Theorem 5.6, Steinwart and Scovel (2007). *Let $\mathcal{F}$ be a convex set of bounded measurable functions from Z to $\mathbb{R}$ and let $L : \mathcal{F} \times Z \to [0, \infty)$ be a convex and line-continuous loss function. For a probability measure P on Z we define*

$$\mathcal{G} := \{L \circ f - L \circ f_{P,\mathcal{F}} : f \in \mathcal{F}\}.$$

Suppose that there are constants $c \geq 0, 0 < \alpha < 1, \delta \geq 0$ and $B > 0$ with $E_P g^2 \leq c(E_P g)^\alpha + \delta$ and $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. Furthermore, assume that $\mathcal{G}$ is separable with respect to $\|\cdot\|_\infty$ and that there are constants $a \geq 1$ and $0 < p < 2$ with

$$\sup_{T \in Z^n} \log N(B^{-1}\mathcal{G}, \epsilon, L_2(T)) \leq a\epsilon^{-p}$$

for all $\epsilon > 0$. Then there exists a constant $c_p > 0$ depending only on p such that for all $n \geq 1$ and all $\tau \geq 1$ we have

$$Pr^*(T \in Z^n : \mathcal{R}_{L,P}(f_{T,\mathcal{F}}) > \mathcal{R}_{L,P}(f_{P,\mathcal{F}}) + c_p \epsilon(n, a, B, c, \delta, \tau)) \leq e^{-\tau},$$

where

$$\epsilon(n, a, B, c, \delta, x) := B^{2p/(4-2\alpha+\alpha p)} c^{(2-p)/(4-2\alpha+\alpha p)} \left(\frac{a}{n}\right)^{2/(4-2\alpha+\alpha p)} + B^{p/2}\delta^{(2-p)/4}\left(\frac{a}{n}\right)^{1/2} + B\left(\frac{a}{n}\right)^{2/(2+p)} + \sqrt{\frac{\delta x}{n}} + \left(\frac{c\tau}{n}\right)^{1/(2-\alpha)} + \frac{B\tau}{n}.$$

In their paper, $f_{P,\mathcal{F}} \in \mathcal{F}$ is a minimizer of $\mathcal{R}_{L,P}(f) = E(L(f,z))$, and $f_{T,\mathcal{F}}$ is similarly defined when T is an empirical measure. To use this theorem, we define $\mathcal{F}, Z, T, \mathcal{G}, f_{T,\mathcal{F}}$ and $f_{P,\mathcal{F}}$ according to our setting. It suffices to consider the subspace of $\mathcal{H}_k$, denoted by $B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)$, as the ball of $\mathcal{H}_k$ of radius $\sqrt{M/\lambda_n}$. Specifically, we let $\mathcal{F}$ be $B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)$ and Z be $\mathcal{X}$. The loss function we consider here is $L_\phi(f) + \lambda_n \|f\|_k^2$ and $\mathcal{G}$ is the function class

$$\mathcal{G}_{\phi,\lambda_n} = \left\{ L_\phi(f) + \lambda_n \|f\|_k^2 - L_\phi(f^*_{\phi,\lambda_n}) - \lambda_n \|f^*_{\phi,\lambda_n}\|_k^2 : f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right) \right\},$$

where $f^*_{\phi,\lambda_n} = \operatorname{argmin}_{f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)} (\lambda_n \|f\|_k^2 + \mathcal{R}_\phi(f))$. $f_{P,\mathcal{F}}$ and $f_{T,\mathcal{F}}$ correspond to f^*_{ϕ,λ_n} and $\hat{f}_n$, respectively. Therefore, to apply this theorem, we will show that there are constants $c \geq 0$ and $B > 0$, which can possibly depend on n, such that $E(g^2) \leq cE(g)$ and $\|g\|_\infty \leq B$, $\forall g \in \mathcal{G}_{\phi,\lambda}$. Moreover, there are constants $\tilde{c}$ and $0 < \nu < 2$ with

$$\sup_{P_n} \log N(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2(P_n)) \leq \tilde{c}\epsilon^{-\nu},$$

for all $\epsilon > 0$.

Let C_L denote $\sup\{R/\min(\pi, 1-\pi)\}$, which is finite provided that R is bounded. Since the weighted hinge loss is Lipschitz continuous with respect to f, with Lipschitz constant C_L, and since $\|f\|_\infty \leq \|f\|_k$ given that $k(x,x) \leq 1$, for any $g \in \mathcal{G}_{\phi,\lambda_n}$, we have

$$\begin{aligned} |g| &\leq |L_\phi(f) - L_\phi(f^*_{\phi,\lambda_n})| + \lambda_n \left| \|f\|_k^2 - \|f^*_{\phi,\lambda_n}\|_k^2 \right| \\ &\leq C_L |f(x) - f^*_{\phi,\lambda_n}(x)| + M \\ &\leq 2C_L\sqrt{M}\lambda_n^{-1/2} + M. \end{aligned} \tag{5.3.5}$$

Therefore, we can set $B = 2C_L\sqrt{M}\lambda_n^{-1/2} + M$.

For any $g \in \mathcal{G}_{\phi,\lambda_n}$, we have

$$\begin{aligned} g(f) &\leq |L_\phi(f) - L_\phi(f^*_{\phi,\lambda_n})| + \lambda_n \left| \|f\|_k^2 - \|f^*_{\phi,\lambda_n}\|_k^2 \right| \\ &\leq C_L |f - f^*_{\phi,\lambda_n}| + \lambda_n \|f - f^*_{\phi,\lambda_n}\|_k \|f + f^*_{\phi,\lambda_n}\|_k \\ &= \left(C_L + 2\sqrt{M\lambda_n} \right) \|f - f^*_{\phi,\lambda_n}\|_k. \end{aligned}$$

Squaring both sides and taking expectations yields

$$E(g^2) \leq \left(C_L + 2\sqrt{M\lambda_n} \right)^2 \|f - f^*_{\phi,\lambda_n}\|_k^2. \tag{5.3.6}$$

On the other hand, from the convexity of L_ϕ, we have

$$\begin{aligned} &\frac{1}{2}(L_\phi(f) + \lambda_n \|f\|_k^2 + L_\phi(f^*_{\phi,\lambda_n}) + \lambda_n \|f^*_{\phi,\lambda_n}\|_k^2) \\ \geq & L_\phi\left(\frac{f + f^*_{\phi,\lambda_n}}{2} \right) + \lambda_n \frac{\|f\|_k^2 + \|f^*_{\phi,\lambda_n}\|_k^2}{2} \\ = & L_\phi\left(\frac{f + f^*_{\phi,\lambda_n}}{2} \right) + \lambda_n \left\| \frac{f + f^*_{\phi,\lambda_n}}{2} \right\|_k^2 + \lambda_n \left\| \frac{f - f^*_{\phi,\lambda_n}}{2} \right\|_k^2 \\ \geq & L_\phi\left(f^*_{\phi,\lambda_n} \right) + \lambda_n \left\| f^*_{\phi,\lambda_n} \right\|_k^2 + \lambda_n \left\| \frac{f - f^*_{\phi,\lambda_n}}{2} \right\|_k^2. \end{aligned}$$

Taking expectations on both sides leads to $E(g) \geq \lambda_n \|f - f^*_{\phi,\lambda_n}\|_k^2/2$. Combining this with (5.3.6), we conclude that $E(g^2) \leq cE(g)$, where

$$c = \frac{2}{\lambda_n} \left(C_L + 2\sqrt{M\lambda_n} \right)^2. \tag{5.3.7}$$

To estimate the bound for $N(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2(P_n))$, we first have

$$N(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2(P_n)) = N\left(B^{-1} \left\{ L_\phi(f) + \lambda_n \|f\|_k^2 : f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n} \right) \right\}, \epsilon, L_2(P_n) \right).$$

From the sub-additivity of the entropy,

$$\log N\left(B^{-1}\mathcal{G}_{\phi,\lambda_n}, 2\epsilon, L_2(P_n)\right) \leq \log N\left(B^{-1}\left\{L_\phi(f) : f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\}, \epsilon, L_2(P_n)\right)$$
$$+ \log N\left(\left\{\lambda_n\|f\|_k^2, f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\}, \epsilon, L_2(P_n)\right). \tag{5.3.8}$$

Using the Lipschitz-continuity of the weighted hinge loss, we now have that if $u, u' \in B^{-1}\{L_\phi(f) : f \in B_{\mathcal{H}_k}(\sqrt{M/\lambda_n})\}$ with corresponding $f, f' \in B_{\mathcal{H}_k}(\sqrt{M/\lambda_n})$, then $\|u - u'\|_{L_2(P_n)} \leq B^{-1}C_L\|f - f'\|_{L_2(P_n)}$, and therefore the first term on the right-hand-side of (5.3.8) satisfies

$$\begin{aligned}
&\log N\left(B^{-1}\left\{L_\phi(f) : f \in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\}, \epsilon, L_2(P_n)\right)\\
\leq &\log N\left(B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right), \frac{B\epsilon}{C_L}, L_2(P_n)\right)\\
\leq &\log N\left(B_{\mathcal{H}_k}, \frac{B\epsilon}{C_L\sqrt{M/\lambda_n}}, L_2(P_n)\right)\\
\leq &\log N\left(B_{\mathcal{H}_k}, 2\epsilon, L_2(P_n)\right).
\end{aligned}$$

The last inequality follows because $B/C_L\sqrt{M/\lambda_n} \geq 2$. It is trivial to see that for the second term on the right hand side of (5.3.8),

$$\log N\left(\left\{\lambda_n\|f\|_k^2, f \in B\left(\sqrt{M/\lambda_n}\right)\right\}, \epsilon, L_2(P_n)\right) \leq \log\left(\frac{M}{B\epsilon}\right).$$

Thus,

$$\log N\left(B^{-1}\mathcal{G}_{\phi,\lambda_n}, 2\epsilon, L_2(P_n)\right) \leq \log N\left(B_{\mathcal{H}_k}, 2\epsilon, L_2(P_n)\right) + \log\left(\frac{M}{B\epsilon}\right).$$

Using (3.3.5) and a given choice for B, we obtain for all $\sigma_n > 0$, $0 < \nu < 2$, $\delta > 0$, $\epsilon > 0$,

$$\sup_{P_n} \log N\left(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2(P_n)\right) \le c_2 \sigma_n^{(1-\nu/2)(1+\delta)d} \epsilon^{-\nu},$$

where c_2 depends on ν, δ and d.

Consequently, from Theorem 5.6 in Steinwart and Scovel (2007), there exists a constant $c_\nu > 0$ depending only on ν such that for all $n \ge 1$ and all $\tau \ge 1$, we have the bound for the first term

$$P^*\left(\lambda_n \|\hat{f}_n\|_k^2 + \mathcal{R}_\phi(\hat{f}_n) > \inf_{f \in \mathcal{H}_k}(\lambda_n \|f\|_k^2 + \mathcal{R}_\phi(f)) + c_\nu \epsilon(n, \tilde{c}, B, c, \tau)\right) \le e^{-\tau},$$

where

$$\epsilon(n, \tilde{c}, B, c, \tau) = \left(B + B^{\frac{2\nu}{2+\nu}} c^{\frac{2-\nu}{2+\nu}}\right)\left(\frac{\tilde{c}}{n}\right)^{\frac{2}{2+\nu}} + (B + c)\frac{\tau}{n}.$$

With B and c as defined in (5.3.5) and (5.3.7), i.e., $\tilde{c} = c_2 \sigma_n^{(1-\nu/2)(1+\delta)d}$ and $\sigma_n = -\lambda_n^{1/(q+1)d}$, we obtain

$$\epsilon(n, \tilde{c}, B, c, \tau) = C_1 \left(\frac{1}{\lambda_n}\right)^{\frac{2}{2+\nu} + \frac{(2-\nu)(1+\delta)}{(2+\nu)(1+q)}} \left(\frac{1}{n}\right)^{\frac{2}{2+\nu}} + C_2 \left(\frac{1}{\lambda_n}\right)^{\frac{q}{q+1}} \frac{\tau}{n}, \tag{5.3.9}$$

where C_1 and C_2 are constants depending on ν, δ, d, M and π. We complete the proof of Theorem 3.3.4 by plugging (5.3.4) and (5.3.9) into (5.3.3).

Proof of Theorem 3.3.5

We apply Theorem 4.3 in Blanchard et al. (2008) on the scaled loss function $\tilde{L}_\phi(f) = L_\phi(f)/C_L$ to obtain the rates in Theorem 3.3.5. Without loss of generality, we can assume that the Bayes classifier $f^* \in \mathcal{H}_k$, since we can always find $g \in \mathcal{H}_k$ such that $\mathcal{R}_\phi(g) = \mathcal{R}_\phi(f^*) = \mathcal{R}_\phi^*$, provided that $\mathcal{H}_k$ is dense in $C(\mathcal{X})$. Let $\mathcal{S}$ be a countable and dense subset of $\mathbb{R}^+$, and let $B_{\mathcal{H}_k}(S)$ denote the ball of $\mathcal{H}_k$ of radius S. Then

$B_{\mathcal{H}_k}(S), S \in \mathcal{S}$ is a countable collection of classes of functions. We can then use Theorem 4.3 in Blanchard et al. (2008) after we verify the following conditions (H1)–(H4):

(H1) $\forall S \in \mathcal{S}, \forall f \in B_{\mathcal{H}_k}(S), \|\tilde{L}_\phi(f)\|_\infty \le b_S, b_S = 1 + S$*;*

(H2) $\forall f, f' \in \mathcal{H}_k,\ Var(\tilde{L}_\phi(f) - \tilde{L}_\phi(f')) \le d^2(f, f'),\ d(f, f') = \|f - f'\|_{L_2(P)}$*;*

(H3) $\forall S \in \mathcal{S}, \forall f \in B_{\mathcal{H}_k}(S), d^2(f, f^*) \le C_S E(\tilde{L}_\phi(f) - \tilde{L}_\phi(f^*)), C_S = 2(S/\eta_0 + 1/\eta_1)$*;*

(H4) Let

$$\xi(x) = \int_0^x \sqrt{\log N(B_{\mathcal{H}_k}, \epsilon, L_2(P_n))}d\epsilon.$$

We have

$$E\Big[\sup_{\substack{f \in B_{\mathcal{H}_k}(S) \\ d^2(f,f') \le r}} (P - P_n)(\tilde{L}_\phi(f) - \tilde{L}_\phi(f'))\Big] \le \inf_{\vartheta > 0}\left\{4\vartheta - \frac{12}{\sqrt{n}}\xi(\vartheta) + \frac{12}{\sqrt{n}}\xi\left(\frac{\sqrt{r}}{\sqrt{2}S}\right)\right\}$$
$$= \psi_S(r).$$

$\psi_S, S \in \mathcal{S}$*, is a sequence of sub-root functions, that is,* ψ_S *is non-negative, non-decreasing, and* $\psi_S(r)/\sqrt{r}$ *is non-increasing for* $r > 0$*. Denote* x_* *as the solution of the equation* $\xi(x) = \sqrt{n}x^2$*. If* r_S^* *denotes the solution of* $\psi_S(r) = r/C_S$*, then*

$$r_S^* \le \inf_{\vartheta > 0} C_S\{4\vartheta - 12\xi(\vartheta)/\sqrt{n}\} + c^2 C_S^2 x_*^2.$$

Under these conditions, we define for $n \in \mathbb{N}$ the following quantity:

$$\gamma_n = \inf_{\vartheta > 0}\left\{4\vartheta - \frac{12}{\sqrt{n}}\xi(\vartheta) + x_*^2(n)\right\}.$$

Given $\mathcal{H}_k$ is associated with the Gaussian kernel, we can show that $\xi(x) \preceq \epsilon^{1-\nu}$ for any $0 < \nu < 2$. Thus, $\gamma_n \preceq \max(n^{-1/2\nu}, n^{-1/(\nu+1)})$. By the choice of $\lambda_n = O(n^{-1/(\nu+1)})$

for any $\nu \in (0,1)$, this satisfies

$$\lambda_n \geq c\left(\gamma_n + \eta_1^{-1}\frac{\log(\tau^{-1}\log n) \vee 1}{n}\right).$$

Therefore, according to Theorem 4.3 in Blanchard et al. (2008), the following bound holds with probability at least $1-\tau$, where $\tau > 0$ is a fixed real number:

$$E(\tilde{L}_\phi(\hat{f}_n)) - E(\tilde{L}_\phi(f^*)) \leq 2 \inf_{f \in \mathcal{H}_k} [E(\tilde{L}_\phi(f)) - E(\tilde{L}_\phi(f^*)) + 2\lambda_n \|f\|_k^2] + 4\lambda_n(8 + c\eta_1\eta_0^{-1}).$$

The result does not change after we scale back to the original loss $L_\phi(f)$. We have shown that $\inf_{f\in\mathcal{H}_k}[\mathcal{R}_\phi(f) - \mathcal{R}_\phi(f^*) + 2\lambda_n\|f\|_k^2] = O(\lambda_n^{q/(q+1)})$ in the proof of Theorem 3.3.4. Thus

$$\mathcal{R}(\hat{f}_n) - \mathcal{R}^* = O_p(\lambda_n^{q/(q+1)}) = O_p\left(n^{-\frac{1}{\nu+1}\frac{q}{q+1}}\right).$$

The remainder of the proof is to verify conditions (H1)–(H4).

For condition (H1), $\|\tilde{L}_\phi(f)\|_\infty \leq \sup\{R/(A\pi + (1-A)/2)\}(1+S)/C_L \leq 1+S$, $\|f\|_k \leq S$.

For condition (H2), let $d(f,f') = \|f - f'\|_{L_2(P)}$. $L_\phi(f)$ is a Lipschitz function with respect to f with Lipschitz constant C_L. Then $\tilde{L}_\phi(f) - \tilde{L}_\phi(f') \leq |f(x) - f'(x)|$. Hence (H2) is easily satisfied.

For condition (H3), the proof is similar to Lemma 6.4 of Blanchard et al. (2008) with $C_S = 2(S/\eta_1 + 1/\eta_0)$, where η_0 and η_1 are as defined in Assumptions (A1) and (A2) of Section 3.5.

For condition (H4), we introduce the notation for Rademacher averages: let $\varepsilon_1, \ldots, \varepsilon_n$ be n i.i.d Rademacher random variables, independent of $(X_i, A_i, R_i), i = 1, \ldots, n$. For any measurable real-valued function f, the Rademacher average is defined as $\mathcal{L}_n f = n^{-1}\sum_{i=1}^n \varepsilon_i f(X_i)$. Also let $\mathcal{L}_n(\mathcal{F})$ be the empirical Rademacher complexity of function class $\mathcal{F}$, $\mathcal{L}_n\mathcal{F} = \sup_{f\in\mathcal{F}} \mathcal{L}_n f$.

First we have from Lemma 6.7 of Blanchard et al. (2008) that for $f' \in \mathcal{H}_k$,

$$E\left[\sup_{f\in\mathcal{H}_k}(P-P_n)(\tilde{L}_\phi(f)-\tilde{L}_\phi(f'))\right] \leq 4E\left[\mathcal{L}_n\{f-f', f\in\mathcal{H}_k\}\right].$$

Thus for the set $\{f\in\mathcal{H}_k : \|f\|_k \leq S, d^2(f,f')\leq r\}$ and $f'\in B_{\mathcal{H}_k}(S)$,

$$E\left[\sup_{f\in\mathcal{H}_k}(P-P_n)(\tilde{L}_\phi(f)-\tilde{L}_\phi(f'))\right] \leq 4E\left[\mathcal{L}_n\{f-f', f\in\mathcal{H}_k : \|f\|_k\leq S, d^2(f,f')\leq r\}\right],$$

the right-hand-side of which is equivalent to $4E\left[\mathcal{L}_n\{f, f\in\mathcal{H}_k : \|f\|_k \leq 2S, \|f\|^2_{L_2(P_n)}\leq 2r\}\right]$. Now we proceed to show that

$$\begin{aligned}&E\mathcal{L}_n\left\{f\in\mathcal{H}_k : \|f\|_k\leq 2S, \|f\|^2_{L_2(P_n)}\leq 2r\right\}\\ &\leq \inf_{\vartheta>0}\left\{4\vartheta+\frac{12}{\sqrt{n}}\int_{\vartheta}^{\frac{\sqrt{r}}{\sqrt{2}S}}\sqrt{\log N(B_{\mathcal{H}},\epsilon,L_2(P_n))}d\epsilon\right\} = \psi_S(r),\end{aligned}$$

by slightly modifying the procedure in obtaining Dudley's Entropy Integral for Rademacher complexity of sets of functions. For $j\geq 0$, let $r_j = \sqrt{2r}2^{-j}$ and T_j be a r_j-cover of $B_{\mathcal{H}_k}(2S)$ with respect to the $L_2(P_n)$-norm. For each $f\in B_{\mathcal{H}_k}(2S)$, we can find an $\tilde{f}_j\in T_j$, such that $\|f-\tilde{f}_j\|_{L_2(P_n)}\leq r_j$. For any N, we express f as $f = f-\tilde{f}_N+\sum_{j=1}^N(\tilde{f}_j-\tilde{f}_{j-1})$, where $\tilde{f}_0=0$. Note $\tilde{f}_0=0$ is an r_0-approximation of f. Hence,

$$\begin{aligned}&\mathcal{L}_n(B_{\mathcal{H}_k}(2S)) = E\left[\sup_{f\in B_{\mathcal{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\varepsilon_i\left(f(X_i)-\tilde{f}_N(X_i)+\sum_{j=1}^N(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i))\right)\right]\\ &\leq E\left[\sup_{f\in B_{\mathcal{H}_k}(2S)}\|\varepsilon\|_{L_2(P_n)}\left\|f-\tilde{f}_N\right\|_{L_2(P_n)}\right]+\sum_{j=1}^N E\left[\sup_{f\in B_{\mathcal{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\left(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i)\right)\right]\\ &\leq r_N+\sum_{j=1}^N E\left[\sup_{f\in B_{\mathcal{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\left(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i)\right)\right].\end{aligned}$$

Note that

$$\left\|\tilde{f}_j - \tilde{f}_{j-1}\right\|_{L_2(P_n)}^2 \leq \left(\left\|\tilde{f}_j - f\right\|_{L_2(P_n)} + \left\|f - \tilde{f}_{j-1}\right\|_{L_2(P_n)}\right)^2 \leq (r_j + r_{j-1})^2 = 9r_j^2.$$

We therefore have

$$\begin{aligned}
\mathcal{L}_n(B_{\mathcal{H}_k}(2S)) &\leq r_N + \sum_{j=1}^{N} 3r_j \sqrt{\frac{2\log(|T_j||T_{j-1}|)}{n}} \\
&\leq r_N + 12\sum_{j=1}^{N}(r_j - r_{j+1})\sqrt{\frac{\log N(B_{\mathcal{H}_k}(2S), r_j, L_2(P_n))}{n}} \\
&\leq r_N + 12\int_{r_{N+1}}^{\sqrt{r}/\sqrt{2}S} \sqrt{\frac{\log N(B_{\mathcal{H}_k}, \epsilon, L_2(P_n))}{n}} d\epsilon.
\end{aligned}$$

For any $\vartheta > 0$, we can choose $N = \sup\{j : r_j > 2\vartheta\}$. Therefore, $\vartheta < r_{N+1} < 2\vartheta$, and $r_N < 4\vartheta$. We therefore conclude that

$$\begin{aligned}
\mathcal{L}_n(B_{\mathcal{H}_k}(2S)) &\leq \inf_{\vartheta > 0}\left\{4\vartheta + 12\int_{\vartheta}^{\sqrt{r}/\sqrt{2}S} \sqrt{\frac{\log N(B_{\mathcal{H}_k}, \epsilon, L_2(P_n))}{n}} d\epsilon\right\} \\
&= \inf_{\vartheta > 0}\left\{4\vartheta - \frac{12}{\sqrt{n}}\xi(\vartheta) + \frac{12}{\sqrt{n}}\xi\left(\frac{\sqrt{r}}{\sqrt{2}S}\right)\right\} = \psi_S(r).
\end{aligned}$$

The function ψ_S is sub-root because $\log N(B_{\mathcal{H}_k}, \epsilon, L_2(P_n))$ is a decreasing function of ϵ.

To show the upperbound of r^*, let $t_S^* = c^2C_S^2x_*^2$. Then $\sqrt{t_S^*}/\sqrt{2}S = cC_Sx_*/\sqrt{2}S$, $C_S/S \geq 1$. Assuming that $c \geq 2$, we have $\sqrt{t_S^*}/\sqrt{2}S \geq x_*$. Since $x^{-1}\xi(x)$ is a decreasing function, it follows that

$$\xi\left(\frac{\sqrt{t_S^*}}{\sqrt{2}S}\right) \leq c\frac{C_S}{\sqrt{2}S}\xi(x_*) = \frac{\sqrt{n}}{cSC_S}t_S^*.$$

Therefore, by selecting an appropriate constant c,

$$\psi_S(t_S^*) \leq \inf_{\vartheta > 0}\left\{4\vartheta - \frac{12}{\sqrt{n}}\xi(\vartheta)\right\} + \frac{12}{cSC_S}t_S^* \leq \frac{C_S \inf_{\vartheta>0}\{4\vartheta - 12/\sqrt{n}\xi(\vartheta)\} + t_S^*}{C_S}.$$

The desired result follows from the property of sub-root functions, which states that if $\psi : [0, \infty) \to [0, \infty)$ is a sub-root function, then the unique positive solution of $\psi(r) = r$, denoted by r^*, exists, and for all $r > 0$, $r \geq \psi(r)$ if and only if $r^* \leq r$ (Bartlett et al., 2005).

Appendix 2: Chapter 4 Proofs

Proof of Theorem 4.3.2

Proof. At stage T, $\tilde{\mathcal{R}}_T(h_T) = \mathcal{R}^*_T(h_T)$. It is not difficult to verify $\hat{V}_{T,n} \to V^*_T$ by applying the consistency results under the single stage setup, if $\lambda_{T,n} \to 0, n_T\lambda_{T,n} \to \infty$ and f^*_T belongs to the closure of $\limsup_n \mathcal{H}_{k_T}$, where $d^*_T = sign(f^*_T)$.

$\hat{f}_{T-1,n}(h_{T-1})$ is obtained via minimizing

$$\mathbb{E}\left[\frac{(R_{T-1} + \hat{V}_{T,n}(\hat{f}_{T,n}))\phi(A_{T-1}f_{T-1}(h_{T-1}))}{A_{T-1}\pi_{T-1} + (1 - A_{T-1})/2}\right| H_{T-1} = h_{T-1},$$
$$A_T = sign(\hat{f}_{T,n}(H_T))\Big] + \lambda_{T-1,n}\|f_{T-1}\|^2.$$

We have

$$\begin{aligned}\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) - \mathcal{R}^*_{T-1}(h_t) &= \hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) - \tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) \\ &\quad + \tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) - \mathcal{R}^*_{T-1}(h_{T-1}),\end{aligned}$$

Using the established consistency results for the single stage setup, $\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) \to \tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n})$ if $\lambda_{T-1,n} \to 0, n_{T-1}\lambda_{T-1,n} \to \infty$ and the minimizer of $\tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n})$

belongs to the closure of $\limsup_n \mathcal{H}_{k_{T-1}}$. We also have

$$\begin{aligned}
&\tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) \\
=&\min_{f_{T-1}} E\left[\frac{(R_{T-1} + V_T^*(H_T) - V_T^*(H_T) + \hat{V}_{T,n}(\hat{f}_{T,n}))I(A_{T-1} \neq sign(f_{T-1}(h_{T-1})))}{A_{T-1}\pi_{T-1} + (1 - A_{T-1})/2}\right| \\
&\qquad\qquad H_{T-1} = h_{T-1}, A_T = sign(\hat{f}_{T,n}(H_T))\Big] \\
\to&\mathcal{R}_{T-1}^*(h_{T-1}),
\end{aligned}$$

since $\hat{V}_{T,n}(\hat{f}_{T,n}) \to V_T^*(h_T)$. Therefore, $\hat{V}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) \to V_{T-1}^*(h_{T-1})$. Repeated arguments lead to the consistency results, that is, $\hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) \to V_t^*(h_t)$, in probability, for all $t = 1, \ldots, T$.

Proof of Theorem 4.3.3

Directly applying the results from the single-decision setup, we obtain that at stage T, if the distribution of (H_T, A_T, R_T) satisfies condition (4.3.5) with noise exponent $q_T > 0$, there exists a constant C_T, depending on ν, δ, p_T and π_T, such that for all $\tau \geq 1$ and $\sigma_{T,n} = \lambda_{T,n}^{-1/(q_T+1)p_T}$,

$$Pr^*(\hat{\mathcal{R}}_{T,n}(f_{T,n}) \leq \mathcal{R}_T^*(h_T) + \epsilon_T) \geq 1 - e^{-\tau},$$

or equivalently,

$$Pr^*(\hat{V}_{T,n}(f_{T,n}) \geq V_T^*(h_T) - \epsilon_T) \geq 1 - e^{-\tau}.$$

Proceeding to stage $T-1$, if the geometric noise exponent condition holds for the distribution $(H_{T-1}, A_{T-1}, R_{T-1})$, there exists a constant C_{T-1} (depending on ν, δ, p_{T-1} and π_{T-1}) such that for all $\tau \geq 1$ and $\sigma_{T-1,n} = \lambda_{T-1,n}^{-1/(q_{T-1}+1)p_{T-1}}$,

$$Pr^*(\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) \leq \tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) + \epsilon_{T-1}) \geq 1 - e^{-\tau}, \qquad (5.3.10)$$

where $\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n})$ and $\tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n})$ are defined in (4.3.1) and (4.3.2) by letting $t = T - 1$. Moreover,

$$Pr^*(\tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) \leq \mathcal{R}^*_{T-1}(h_{T-1}) + \epsilon_T) \geq 1 - e^{-\tau}, \tag{5.3.11}$$

Combining (5.3.10) and (5.3.11), we obtain that

$$Pr^*(\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) \leq \mathcal{R}^*_{T-1}(h_{T-1}) + \epsilon_{T-1} + \epsilon_T) \geq 1 - 2e^{-\tau}.$$

Note that

$$\begin{aligned}
&V^*_{T-1}(h_{T-1}) - \hat{V}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n})\\
=&E\left[\left. \frac{(R_{T-1} + V^*_T(H_T)) - (R_{T-1} + \hat{V}_{T,n}(\hat{f}_{T,n}))}{A_{T-1}\pi_{T-1} + (1 - A_{T-1})/2} \right| H_{T-1} = h_{T-1}, A_T = \text{sign}(\hat{f}_{T,n}(H_T)) \right]\\
+&\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) - \mathcal{R}^*_{T-1}(h_{T-1}).
\end{aligned}$$

Thus

$$Pr^*(\hat{V}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) \geq V^*_{T-1}(h_{T-1}) - \epsilon_{T-1} - 2\epsilon_T) \geq 1 - 3e^{-\tau}.$$

Repeating the arguments, we obtain that at stage t, if for stages $l, l = t, \ldots, T$, $\sigma_{l,n} = \lambda_{l,n}^{-1/(q_l+1)p_l}$,

$$Pr^*\left(\hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) \geq V^*_t(h_t) - \sum_{l=t}^{T} 2^{l-t}\epsilon_l\right) \geq 1 - \sum_{l=t}^{T} 2^{l-t}e^{-\tau}.$$

Proof of Theorem 4.3.4

We can apply the conclusion from the single stage decision problem to stage T. For any $\nu \in (0,1)$ and $q_T \in (0,\infty)$, let $\lambda_{T,n} = O(n_T^{-1/(\nu+1)})$ and $\sigma_{T,n} = \lambda_{T,n}^{-1/(q_T+1)p_T}$. Then

$$\hat{\mathcal{R}}_{T,n}(\hat{f}_{T,n}) - \mathcal{R}_T^*(h_T) = O_p\left(n_T^{-\frac{1}{\nu}\frac{q_T}{q_T+1}}\right),$$

indicating

$$V_T^*(h_T) - \hat{V}_{T,n}(\hat{f}_{T,n}) = O_p\left(n_T^{-\frac{1}{\nu}\frac{q_T}{q_T+1}}\right).$$

For stage $T-1$, the analysis is restricted to the subset of patients whose assigned treatments are the same as their estimates in stage T. With the available sample size denoted as n_{T-1}, for any $\nu \in (0,1)$ and $q_{T-1} \in (0,\infty)$, if (A1) and (A2) are satisfied and $\lambda_{T-1,n}$ and $\sigma_{T-1,n}$ are appropriately selected, we have

$$\hat{\mathcal{R}}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) - \tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) = O_p\left(n_{T-1}^{-\frac{1}{\nu}\frac{q_{T-1}}{q_{T-1}+1}}\right).$$

In addition,

$$\tilde{\mathcal{R}}_{T-1}(h_{T-1}, \hat{f}_{T,n}) - \mathcal{R}^*(h_{T-1}) = O_p\left(n_T^{-\frac{1}{\nu}\frac{q_T}{q_T+1}}\right).$$

Thus,

$$V_{T-1}^*(h_{T-1}) - \hat{V}_{T-1,n}(\hat{f}_{T-1,n}, \hat{f}_{T,n}) = O_p\left(n_{T-1}^{-\frac{1}{\nu}\frac{q_{T-1}}{q_{T-1}+1}}\right) + O_p\left(n_T^{-\frac{1}{\nu}\frac{q_T}{q_T+1}}\right).$$

Recycling arguments, we have

$$V_t^*(h_t) - \hat{V}_{t,n}(\hat{f}_{t,n}, \ldots, \hat{f}_{T,n}) = \sum_{l=t}^{T} O_p\left(n_l^{-\frac{1}{\nu}\frac{q_l}{q_l+1}}\right),$$

and the desired result follows.

Bibliography

Bartlett, P. L., Bousquet, O., and Mendelson, S. (2005), "Local Rademacher Complexities," *The Annals of Statistics*, 33, 1497–1537.

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2006), "Convexity, Classification, and Risk Bounds," *J. of American Statistical Association*, 101, 138–156.

Bellman, R. (1957), *Dynamic Programming*, Princeton: Princeton Univeristy Press.

Blanchard, G., Bousquet, O., and Massart, P. (2008), "Statistical Performance of Support Vector Machines," *Annals of Statistics*, 36, 489–531.

Blatt, D., Murphy, S. A., and Zhu, J. (2004), "A-learning for approximate planning," Unpublished Manuscript.

Bradley, P. S. and Mangasarian, O. L. (1998), "Feature Selection via Concave Minimization and Support Vector Machines," in *Proc. 15th International Conf. on Machine Learning*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Buzdar, A. U. (2009), "Role of Biologic Therapy and Chemotherapy in Hormone Receptor and HER2-Positive Breast Cancer," *Annals of Oncology*, 20, 993–999.

Cai, T., Tian, L., Uno, H., and Solomon, S. D. (2010), "Calibrating parametric subject-specific risk estimation," *Biometrika*, 97, 389–404.

Chakraborty, B., Murphy, S. A., and Strecher, V. (2009), "Inference for non-regular parameters in optimal dynamic treatment regimes," *Stat Methods Med Res*, 19, 317–43.

Cortes, C. and Vapnik, V. (1995), "Support-Vector Networks," in *Machine Learning*, pp. 273–297.

Craigmile, P. F., Kim, N., Fernandez, S. A., and Bonsu, B. K. (2007), "Modeling and detection of respiratory-related outbreak signatures," *BMC Medical Informatics and Decision Making*, 7.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley, 2nd ed.

Crits-Christoph, P., Siqueland, L., Blaine, J., Frank, A., Luborsky, L., Onken, L. S., Muenz, L. R., Thase, M. E., Weiss, R. D., Gastfriend, D. R., Woody, G. E., Barber, J. P., Butler, S. F., Daley, D., Salloum, I., Bishop, S., Najavits, L. M., Lis, J., Mercer, D., Griffin, M. L., Moras, K., and Beck, A. T. (1999), "Psychosocial Treatments for Cocaine Dependence," *Arch Gen Psychiatry*, 56, 493–502.

Cucala, L., Demattei, C., Lopes, P., and Ribeiro, A. (2009), "Spatial scan statistics for case event data based on connected components," Unpublished Manuscript.

Daley, D. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes*, New York: Springer, 2nd ed.

Dawson, R. and Lavori, P. (2004), "Placebo-free designs for evaluating new mental health treatments: the use of adaptive treatment strategies," *Statistics in Medicine*, 23, 3249–3262.

Dhanasekaran, S., Barrette, T., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. (2001), "Delineation of prognostic biomarkers in prostate cancer," *Nature*, 412, 822–826.

Diggle, P. J., Kaimi, I., and Abellana, R. (1985), "A kernel method for smoothing point process data," *Applied Statistics*, 34, 138–147.

Diggle, P. J., Rowlingson, B., and li Su, T. (2005), "Point process methodology for on-line spatio-temporal disease surveillance," *Environmetrics*, 16, 423–434.

Eagle, K. A., Lim, M. J., Dabbous, O. H., Pieper, K. S., Goldberg, R. J., de Werf, F. V., Goodman, S. G., Granger, C. B., Steg, P. G., Joel M. Gore, M., Budaj, A., Avezum, A., Flather, M. D., Fox, K. A. A., and GRACE Investigators, . (2004), "A Validated Prediction Model for All Forms of Acute Coronary Syndrome: Estimating the Risk of 6-Month Postdischarge Death in an International Registry," *J. Am. Med. Assoc.*, 291, 2727–33.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998), "Cluster analysis and display of genome-wide expression patterns," *Proc Natl Acad Sci U S A*, 95, 14863–14868.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman & Hall.

Flume, P. A., O'Sullivan, B. P., Goss, C. H., Peter J. Mogayzel, J., Willey-Courand, D. B., Bujan, J., Finder, J., Lester, M., Quittell, L., Rosenblatt, R., Vender, R. L., Hazle, L., Sabadosa, K., and Marshall, B. (2007), "Cystic Fibrosis Pulmonary Guidelines: Chronic Medications for Maintenance of Lung Health," *Am. J. Respir. Crit. Care Med.*, 176, 957–969.

Fukuoka, M., Wu, Y.-L., Thongprasert, S., Sunpaweravong, P., Leong, S.-S., Sriuranpong, V., Chao, T.-Y., Nakagawa, K., Chu, D.-T., Saijo, N., Duffield, E. L., Rukazenkov, Y., Speake, G., Jiang, H., Armour, A. A., To, K.-F., Yang, J. C.-H., and Mok, T. S. (2011), "Biomarker Analyses and Final Overall Survival Results From a Phase III, Randomized, Open-Label, First-Line Study of Gefitinib Versus Carboplatin/Paclitaxel in Clinically Selected Patients With Advanced NonSmall-Cell Lung Cancer in Asia (IPASS)," *J Clin Oncol.*, 29, 2866–2874.

Gangnon, R. E. and Clayton, M. K. (2004), "Likelihood-based tests for localized spatial clustering of disease," *Environmetrics*, 15, 797–810.

Goldberg, Y. and Kosorok, M. R. (2012), "Q-Learning with Censored Data," *The Annals of Statistics*, in Press.

Grünwald, V. and Hidalgo, M. (2003), "Developing Inhibitors of the Epidermal Growth Factor Receptor for Cancer Treatment," *J Natl Cancer Inst*, 95, 851–867.

Hamburg, M. and Collins, F. (2010), "The path to personalized medicine," *N Engl J Med.*, 363, 301–304.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009), *The Elements of Statistical Learning*, New York: Springer-Verlag New York, Inc., 2nd ed.

Hutwagner, L., Maloney, E., Bean, N., L, L. S., and Martin, S. (1997), "Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks," *Emerging Infectious Diseases*, 3, 395–400.

Insel, T. R. (2009), "Translating scientific opportunity into public health impact: a strategic plan for research on mental illness," *Archives of General Psychiatry*, 66, 128–133.

Ishigooka, J., Murasaki, M., and Miura, S. (2000), "Olanzapine optimal dose: results of an open-label multicenter study in schizophrenic patients. Olanzapine Late-Phase II Study Group," *Psychiatry Clin Neurosci.*, 54, 467–478.

Jung, I., Kulldorff, M., and Klassen, A. C. (2007), "A spatial scan statistic for ordinal data," *Statistics in Medicine*, 26, 1594–1507.

Karr, A. R., Banks, D., Datta, G., Lynch, J., and Vera, F. (2009), "Bayesian methods in Syndromic Surveillance: CAR Models and computational Implementation," *Project Description*.

Keller, M. B., Mccullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff, C. B., Russell, J. M., Thase, M. E., Trivedi, M. H., and Zajecka, J. (2000), "A Comparison of Nefazodone, The Cognitive Behavioral-Analysis System of Psychotherapy, and Their Combination for the Treatment of Chronic Depression," *The New England Journal of Medicine*, 342, 1462–70.

Kleinman, K., Lazarus, R., and Platt, R. (2004), "A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism," *American Journal of Epidemiology*, 159, 217–224.

Kulldorff, M. (1997), "A spatial scan statistic," *Communications in Statistics - Theory and Methods*, 26, 1481–1496.

— (2001), "Prospective time periodic geographical disease surveillance using a scan statistic," *J. of Royal Statistical Society Series A*, 164, 61–72.

Kulldorff, M., Athas, W., Feurer, E., Miller, B., and Key, C. (1998), "Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico," *American Journal of Public Health*, 88, 1377–1380.

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R., and Mostashari, F. (2005), "A space-time permutation scan statistic for disease outbreak detection," *PLOS Medicine*, 2, 1–9.

Kulldorff, M., Mostashari, F., Duczmal, L., Yih, K., Kleinman, K., and Platt, R. (2007), "Multivariate Scan Statistics for Disease Surveillance," *Statistics in Medicine*, 26, 1824–1833.

Kulldorff, M. and Nagarwalla, N. (1995), "Spatial Disease Clusters: Detection and Inference," *Statistics in Medicine*, 14, 799–810.

Laber, E. B. and Murphy, S. A. (2011), "Adaptive Confidence Intervals for the Test Error in Classification," *Journal of the American Statistical Association*, 106, 904–913.

Laber, E. B., Qian, M., Lizotte, D., Pelham, W. E., and Murphy, S. (2011), "Statistical Inference in Dynamic Treatment Regimes," *Revision of Univ. of Michigan, Statistics Department Technical Report 506.*

Lavori, P. W. and Dawson, R. (2000), "A design for testing clinical strategies: biased adaptive within-subject randomization," *Journal Of The Royal Statistical Society Series A*, 163, 29–38.

— (2004), "Dynamic treatment regimes: practical design considerations," *Clinical Trials*, 1, 9–20.

Lee, Y., Lin, Y., and Wahba, G. (2004), "Multicategory Support Vector Machines, theory, and application to the classification of microarray data and satellite radiance data," *Journal of the American Statistical Association*, 99, 67–81.

Lewis, M., Pavlin, J., Mansfield, J., O'Brien, S., Boomsma, L., Elbert, Y., and Kelley, P. (2002), "Disease outbreak detection system using syndromic data in the greater Washington DC area," *American Journal of Preventive Medicine*, 13, 180–186.

Liu, Y., Helen Zhang, H., Park, C., and Ahn, J. (2007), "Support vector machines with adaptive L_q penalty," *Comput. Stat. Data Anal.*, 51, 6380–94.

Lugosi, G. and Vayatis, N. (2004), "On the Bayes-risk consistency of regularized boosting methods," *The Annals of Statistics*, 32, 30–55.

Lunceford, J. K., Davidian, M., and Tsiatis, A. A. (2002), "Estimation of Survival Distributions of Treatment Policies in Two-Stage Randomization Designs in Clinical Trials," *Biometrics*, 58, 48–57.

Marlowe, D. B., Festinger, D. S., Dugosh, K. L., Lee, P. A., and Benasutti, K. M. (2007), "Adapting Judicial Supervision to the Risk Level of Drug Offenders: Discharge and 6-month Outcomes from a Prospective Matching Study," *Drug and Alcohol Dependence*, 88(Suppl 2), S4–S13.

Møller, J. and Waagepetersen, R. P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, London: Chapman & Hall/CRC.

Moodie, E. E. M., Platt, R. W., and Kramer, M. S. (2009), "Estimating Response-Maximized Decision Rules With Applications to Breastfeeding," *Journal of the American Statistical Association*, 104, 155–165.

Moodie, E. E. M., Richardson, T. S., and Stephens, D. A. (2007), "Demystifying Optimal Dynamic Treatment Regimes," *Biometrics*, 63, 447–455.

Murphy, S. A. (2003), "Optimal Dynamic Treatment Regimes," *Journal of the Royal Statistical Society, Series B*, 65, 331–366.

— (2005a), "An experimental design for the development of adaptive treatment strategies," *Statistics in Medicine*, 24, 1455–1481.

— (2005b), "A Generalization Error for Q-Learning," *Journal of Machine Learning Research*, 6, 1073–1097.

Murphy, S. A., Oslin, D. W., Rush, A. J., Zhu, J., and MCATS (2007), "Methodological Challenges in Constructing Effective Treatment Sequences for Chronic Psychiatric Disorders," *Neuropsychopharmacology*, 32, 257–262.

Murphy, S. A., van der Laan, M. J., Robins, J. M., and CPPRG (2001), "Marginal Mean Models for Dynamic Regimes," *Journal of the American Statistical Association*, 96, 1410–23.

Neill, D. B., Moore, A., Sabhnani, M., and Daniel, K. (2005), "Detection of emerging space-time clusters," *Proc. 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining.*

Neyman, J. S. (1990), "On the Application of Probability Theory to Agricultural Experiments (Engl. transl. by D.M. Dabrowska and T.P. Speed)," *Statistical Science*, 5, 465–472.

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *J. of American Statistical Association*, 83, 9–27.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. (1987), "A Mark 1 geographical analysis machine for the automated analysis of point data sets," *International Journal of Geographical Information Systems*, 1, 335–358.

Pineau, J., Bellemare, M. G., J., R. A., Ghizaru, A., and Murphy, S. A. (2007), "Constructing evidence-based treatment strategies using methods from computer science," *Drug and Alcohol Dependence*, 88S, S52–S60.

Piper, W. E., Boroto, D. R., Joyce, A. S., McCallum, M., and Azim, H. F. A. (1995), "Pattern of alliance and outcome in short-term individual psychotherapy," *Psychotherapy*, 32, 639–647.

Qian, M. and Murphy, S. A. (2011), "Performance Guarantees for Individualized Treatment Rules," *The Annals of Statistics*, 39, 1180–1210.

Reis, B. and Mandl, K. (2003), "Time series modeling for syndromic surveillance," *BMC Medical Informatics and Decision Making*, 3.

Ren, Z., Davidian, M., George, S. L., Goldberg, R. M., Wright, F. A., Tsiatis, A. A., and Kosorok, M. R. (2012), "Research Methods for Clinical Trials in Personalized Medicine: A Systematic Review," Submitted.

Ripley, B. D. (1977), "Modelling spatial patterns (with discussion)," *J. of Royal Statistical Society Series B*, 39, 172–212.

Robins, J. (1986), "A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect," *Mathematical Modelling*, 7, 1393–1512.

— (1997), "Causal inference from complex longitudinal data," *Lect. Notes Statist.*, 120, 69–117.

Robins, J. M. (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," in *In Proceedings of the Second Seattle Symposium on Biostatistics*, Springer, pp. 189–326.

Rosenbaum, P. R. and Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., and et al (2002), "The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma," *New England J. of Medicine*, 1937–47.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688–701.

— (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58.

— (1986), “Comment:“Which Ifs Have Causal Answers”,” *Journal of the American Statistical Association*, 81, 961–962.

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., Thase, M. E., Nierenberg, A. A., Quitkin, F. M., Kashner, T. M., Kupfer, D. J., Rosenbaum, J. F., Alpert, J., Stewart, J. W., McGrath, P. J., Biggs, M. M., Shores-Wilson, K., Lebowitz, B. D., Ritz, L., and Niederehe, G. (2004), “Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design,” *Controlled Clinical Trials*, 25, 119–142.

Sargent, D. J., Conley, B. A., Allegra, C., and Collette, L. (2005), “Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials,” *Journal of Clinical Oncology*, 32, 2020–27.

Schneider, L., Tariot, P., Lyketsos, C., Dagerman, K., Davis, K., Davis, S., Hsiao, J., Jeste, D., Katz, I., Olin, J., Pollock, B., Rabins, P., Rosenheck, R., Small, G., Lebowitz, B., and Lieberman, J. (2001), “National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology,” *American Journal of Geriatric Psychiatry*, 9, 346–360.

Schoenberg, F. P. (1999), “Transforming spatial point processes into Poisson processes,” *Stochastic Processes and their Applications*, 81, 155–164.

— (2003), “Multi-Dimensional Residual Analysis of Point Process Models for Earthquake Occurrences,” *Journal of the American Statistical Association*, 98, 789–795.

— (2004), “Consistent parametric estimation of the intensity of a spatial-temporal point process,” *J. of Statistical Planning and Inference*, 128, 79–93.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., , and Davidian, M. (2012), “Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes,” Submitted.

Socinski, M. and Stinchcombe, T. (2007), “Duration of first-line chemotherapy in advanced non small-cell lung cancer: less is more in the era of effective subsequent therapies,” *J Clin Oncol.*, 25, 5155–5157.

Steinwart, I. (2005), “Consistency of Support Vector Machines and Other Regularized Kernel Classifiers,” *IEEE Transactions on Information Theory*, 51, 128–142.

Steinwart, I. and Scovel, C. (2007), “Fast Rates for Support Vector Machines using Gaussian Kernels,” *The Annals of Statistics*, 35, 575–607.

Strecher, V., McClure, J., Alexander, G., Chakraborty, B., Nair, V., Konkel, J., Greene, S., Collins, L., Carlier, C., Wiese, C., Little, R., Pomerleau, C., and Pomerleau, O. (2008), “Web-based smoking cessation components and tailoring depth: Results of a randomized trial,” *American Journal of Preventive Medicine*, 34, 373–381.

Sutton, R. S. and Barto, A. G. (1998), *Reinforcement Learning I: Introduction*, Cambridge,MA: MIT Press.

Thacker, S. B. and Berkelman, R. L. (1988), "Public health surveillance in the united states," *Epidemiologic Review*, 10, 164–190.

Thall, P. F., Millikan, R. E., and Sung, H. G. (2000), "Evaluating multiple treatment courses in clinical trials," *Statistics in Medicine*, 19, 1011–1028.

Thall, P. F., Sung, H.-G., and Estey, E. H. (2002), "Selecting Therapeutic Strategies Based on Efficacy and Death in Multicourse Clinical Trials," *Journal of the American Statistical Association*, 97, 29–39.

Thall, P. F., Wooten, L. H., Logothetis, C. J., Millikan, R. E., and Tannir, N. M. (2007), "Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring," *Statistics in Medicine*, 26, 4687–4702.

Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002), "Diagnosis of multiple cancer types by shrunken centroids of gene expression," *Proc Natl Acad Sci U S A.*, 99, 6567–6572.

Tsui, F.-C., Espino, J., Dato, V., Gesteland, P., Hutman, J., and Wagner, M. (2003), "Technical Description of RODS: A real-time public health surveillance system," *J. of the American Medical Informatics Association*, 10, 399–408.

Tsybakov, A. B. (2004), "Optimal Aggregation of Classifiers in Statistical Learning," *Annals of Statistics*, 32, 135–166.

van't Veer, L. J. and Bernards, R. (2008), "Enabling Personalized Cancer Medicine through Analysis of Gene-Expression Patterns," *Nature*, 452, 564–570.

Vapnik, V. N. (1995), *The nature of statistical learning theory*, New York: Springer-Verlag New York, Inc.

Wahed, A. S. and Tsiatis, A. A. (2004), "Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials," *Biometrics*, 60, 124–133.

— (2006), "Semi-parametric efficient estimation of the survival distribution for treatment policies in two-stage randomization designs in clinical trials with censored data," *Biometrika*, 93, 147–161.

Wang, L. and Shen, X. (2006), "Multi-category Support vector machines, feature selection, and solution path," *Statistica Sinica*, 16, 617–633.

Watkins, C. J. C. H. (1989), "Learning from delayed rewards," *Ph.D. Thesis, Kings College, Cambridge, U.K.*

Yeung, K. and Ruzzo, W. (2001), "Principal component analysis for clustering gene expression data." *Bioinformatics*, 17, 763–774.

Zhang, T. (2004), "Statistical behavior and consistency of classification methods based on convex risk minimization," *Annals of Statistics*, 32, 56–134.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009), "Reinforcement learning design for cancer clinical trials," *Statistics in Medicine*, 28, 3294–3315.

Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011a), "Reinforcement Learning Strategies for Clinical Trials in Nonsmall Cell Lung Cancer," *Biometrics*, 67, 1422–1433.

Zhao, Y. Q., Zeng, D., Herring, A. H., Ising, A., Waller, A., Richardson, D., and Kosorok, M. R. (2011b), "Detecting disease outbreaks using local spatiotemporal methods," *Biometrics*, 67, 1508–1517.

Zhao, Y. Q., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), "Estimating Individualized Treatment Rules using Outcome Weighted Learning," In Revision (invited).

Zhu, J., Rosset, S., Hastie, T., and Tibshirani, R. (2003), "1-norm Support Vector Machines," in *Neural Information Processing Systems*, p. 16.

Zou, H. and Yuan, M. (2008), "The F_∞-norm Support Vector Machine," *Statistica Sinica*, 18, 379–398.