1-1-2011

Probabilistic models for patient scheduling

Adel Alaeddini
Wayne State University,

Recommended Citation

Alaeddini, Adel, "Probabilistic models for patient scheduling" (2011). *Wayne State University Theses.* Paper 66.

PROBABILISTIC MODELS FOR PATIENT SCHEDULING

by

ADEL ALAEDDINI

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

2011

MAJOR: COMPUTER SCIENCE
(Data Mining)

Approved by:

Advisor Date

DEDICATION

For my family who helped me to make my wishes come true

ACKNOWLEDGEMENTS

I acknowledge Dr. Chandan K. Reddy, my advisor in the Department of Computer Science for his guidance throughout my thesis. Also, I am very thankful to Dr. Farshad Fotouhi in the Department of Computer Science and Dr. Kai Yang (my Ph.D. thesis advisor) in the Department of Industrial Engineering for giving me an opportunity to pursue my Masters degree in Computer Science simultaneously with my Ph.D. in Industrial Engineering. Finally, I kindly appreciate the support of my family and friends.

TABLE OF CONTENTS

Dedication ii

Acknowledgements iii

List of Tables vi

List of Figures vii

CHAPTER 1 INTRODUCTION 1

1.1. Relevant Background 2

1.1.1. Factors Affecting No-Shows and Cancellations 3

1.1.2. Population based Models 3

1.1.3. Individual based Models 4

1.2. Preliminaries 6

1.2.1. Notations Used for the Probability of Disturbance Prediction Model 6

1.2.2. Binomial and Multinomial Logistic Regression 7

1.2.3. Beta and Dirichlet Distributions 8

1.2.4. Bayesian Update of Beta and Dirichlet Distributions 9

1.2.5. Hotelling's T^2 Distribution 11

CHAPTER 2 PREDICTING DISTURBANCES IN APPOINTMENT SCHEDULING THROUGH HYBRID PROBABILISTIC MODELING 13

2.1. The Proposed Algorithm for No-Show Prediction 13

2.2. Experimental Results 16

2.2.1. Data Preprocessing 16

2.2.2. Applying the Proposed Model to a Sample Patient 19

2.2.2.1. Time wise Analysis 22

2.2.2.2. Patient wise Analysis 26

2.2.2.3. Discussion 27

2.3. Generalization of the Proposed Algorithm for No-Show and Cancellation Prediction 28

2.4. Experimental Results.. 31

2.4.1. Data Preprocessing.. 32

2.4.2. Applying the Proposed Model to a Sample Patient.. 33

2.4.2.1. Time wise Analysis 36

2.4.2.2. Patient wise Analysis 40

2.4.2.3. Discussion 42

CHAPTER 3 AN OPTIMIZATION MODEL FOR EFFECTIVE APPOINTMENT SCHEDULING IN THE PRESENCE OF DISTURBANCES 44

3.1. Introduction 44

3.2. The Proposed Optimization Model for Non-Sequential Scheduling 47

3.3. An Optimization Model for Including changing no-show probabilities and walk-in Patients 49

3.3.1. Results on Simulated Data 51

3.4. Generalization of the Proposed Model for Sequential Scheduling 53

3.4.1. Estimation of Demand.. 56

3.4.2. Results on Simulated Data 57

CHAPTER 4 CONCLUSIONS AND FUTURE DIRECTIONS 58

Appendix A -Gaussian Mixture Models (GMM) and Expectation Maximization (EM) Algorithm 59

References 61

Abstract 65

Autobiographical Statement.. 66

LIST OF TABLES

Table 1: Notations used for the proposed prediction model 6

Table 2: Data structure for the weighting factors 19

Table 3: Attendance record of a sample patient........ 20

Table 4: A sample logistic regression model fitted to the dataset 20

Table 5: Bayesian update of Beta distribution parameters 21

Table 6: The result of clustering the clinics based on their no-show, cancellation and show-up probabilities........ 32

Table 7: Data structure and optimal value of the weighting factors 33

Table 8: No-show, cancellation and show-up record of a sample patient 34

Table 9: A sample multiple logistic regression model fitted to the dataset 34

Table 10: Bayesian update of Dirichlet distribution parameters 35

Table 11: Notations used in the proposed scheduling model 46

LIST OF FIGURES

Figure 1: (a) The histogram of clinics no-show probability, (b) The result of clustering the clinics based on their no-show probabilities 18

Figure 2: Applying the proposed model for a sample patient :(a) Real record of attendance and estimated probability of no-show using the proposed model (b) Prior and posterior of Beta distribution for modeling no-show 22

Figure 3: Mean Squared Error (MSE) of the other methods used for comparison 23

Figure 4: Proposed approach performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability 24

Figure 5: Logistic regression performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability 25

Figure 6: ARIMA (1,1,0) performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability 26

Figure 7: Mean squared error (MSE) of different methods used for comparison comparing methods 27

Figure 8: Applying the proposed model for a sample patient: (a) changing parameters during Bayesian update (b) prior distribution (c) posterior distribution 36

Figure 9: Mean Squared Error (MSE) of the studied methods for time-wise analysis 38

Figure 10: The performance of the proposed approach over different patients: estimated versus empirical probability of no-show and cancellation 38

Figure 11: The performance of pure multinomial logistic regression over different patients: estimated versus empirical probability of no-show and cancellation 39

Figure 12: The performance of pure Bayesian updating method over different patients: estimated versus empirical probability of no-show and cancellation 40

Figure 13: Mean Squared Error (MSE) of the studied methods for patient-wise analysis 41

Figure 14: The performance of the proposed approach over different patients: estimated versus empirical probability of no-show and cancellation 41

Figure 15: The performance of pure multinomial logistic regression over different patients: estimated versus empirical probability of no-show and cancellation 42

Figure 16: The performance of pure Bayesian updating method over different patients: estimated versus empirical probability of no-show and cancellation 42

Figure 17: The expected total profit of the proposed non-sequential and the original scheduling model under different no-show probabilities for different types of patients................ 52

Figure 18: The expected total profit of the proposed non-sequential and the original scheduling model under number of demands 53

Figure 19: The expected total profit of the proposed sequential and the original scheduling model across a sequence of 40 patient calls 57

CHAPTER 1 INTRODUCTION

The problem of no-shows and appointments cancellation (individuals who do not arrive for or cancel their scheduled appointments) cause significant disturbance on the smooth operation of almost all scheduling systems [Bech 2005; Moore et al. 2001]. While the reasons for these no-shows, and cancellations might vary from previous experience to personal behaviors, several practitioners and researchers have often neglected this important realistic aspect of the scheduling problem. This thesis, considers the problem of effective scheduling by predicting such disturbances accurately from the historical data available and incorporating them into scheduling using a novel optimization model. Specifically, the applicability and usefulness of the proposed work is demonstrated on healthcare data collected from a medical center. Due to the vast amounts of cost and resources involved in medical healthcare centers, such disturbances can incur losses of hundreds of thousands of dollars yearly [Bech 2005; Hixon et al. 1999; Rust et al. 1995; Barron 1980]. Such disruptions not only cause inconvenience to the hospital management but also has a significant impact on the revenue, cost and resource utilization for almost all the healthcare systems. Hence, accurate prediction of no-show and cancellation probabilities and incorporating them into the scheduling system is a cornerstone for any non-attendance reduction strategy [Cayirli and Veral 2003; Ho and Lau 1992; Cote 1999; Hixon et al. 1999; Moore et al. 2001].

In this research, a hybrid probabilistic model is developed to predict the probability of no-shows and cancellations in real-time using logistic regression and Bayesian inference. In addition, a novel optimization model which can effectively utilize no-show probabilities for scheduling patients is also developed. The proposed prediction model uses both the general social and demographic information of the individuals and their clinical appointments attendance records, and other variables such as the effect of appointment date, and clinic type. In the mean time, the scheduling model considers both scheduled and unscheduled patients (walk-in patients) simultaneously. It also formulates the effect of patients' overflow from one slot to another. In addition, it takes into account the effect of patients' assignment to undesired appointment time on no-show/cancellation probability.

The result of the proposed method can be used to develop more effective appointment scheduling [Chakraborty et al. 2010; Glowacka et al 2009; Gupta and Denton 2008; Hassin and Mendel 2008; Liu et al. 2009]. It can also be used for developing effective strategies such as selective overbooking for reducing the negative effect of disturbances and filling appointment slots while maintaining short waiting times [Laganga and Lawrence 2007; Muthuraman and Lawley 2008; Zeng et al 2010].

The organization of this thesis is as follows: the rest of this chapter discusses the relevant background and preliminaries of this research. Chapter 2 describes the proposed models for predicting disturbances in appointment scheduling and the results of applying the proposed models on data collected from a medical healthcare center. Chapter 3 presents the proposed optimization models for effective appointment scheduling in the presence of disturbances along with two simulated numerical examples. Finally, chapter 4 concludes our work and presents some future extensions of this study.

1.1. Relevant Background

There are wide varieties of techniques that can be used for the estimation of no-show and cancellation probabilities. First, the factors that can affect no-shows and cancellations are briefly discussed. Next, some of the related quantitative methods studied in this domain are presented.

1.1.1. Factors Affecting No-Shows and Cancellations

There have been a few studies that discussed the effect of patients’ personal information such as age, gender, nationality, and population sector on the no-show and cancelation probabilities [Bean and Talaga 1995; and Glowacka 2009]. Some researchers have also investigated the relationship between no-show probability and factors related to the previous (appointments) experience of the person such as number of previous appointments, appointment lead times, waiting times, appointment type, and service quality [Cynthia et al 1995; Garuda et al. 1998; Goldman et al 1982; Dreihera et al. 2008; Lehmann et al. [2007]. A few studies also considered the effect of personal issues such as overslept or forgot, health status, presence of a sick child or relative, and lack of transportation on missing appointments [Campbell et al 2000; Cashman et al. 2004]. This study will consider many of these factors in our proposed model and also consider the effect of personal behavior such as previous appointment-keeping pattern as discussed in [Dove and Karen 1981] in predicting no-shows.

1.1.2. Population based Models

Population based techniques mainly use a variety of methods drawn from statistics and machine learning which can be used for predicting no-shows and cancellations [Dove and Schneider 1981]. These methods use the information from the entire population (dataset) in the form of set factors, in order to estimate the probability of no-show, cancellation and show-up. Logistic regression is one of the most popular statistical methods in this category that is used for binomial and multinomial regression, which can predict the probability of disturbances by fitting numerical or categorical predictor variables in the data to a logit function (Hilbe [2009]). There has been some work using tree-based and rule-based models which create if–then constructs to separate the data into increasingly homogeneous subsets, based on which the desired predictions of disturbances can be made [Glowacka et al. 2009]. The problem with these population based methods is that although they provide a reasonable estimate, they do not differentiate between the behaviors of individual persons, and hence cannot update effectively especially while using small datasets. Another problem with these methods is that once the model has been built adding new data has minor effect on the result especially when the size of initial dataset is much larger compared to the size of the new data. In chapter 2, we compare some of above methodologies with our proposed approach on real-world patient scheduling data.

1.1.3. Individual based Models

Individual based approaches are primarily time series and smoothing methods that are used for predicting the probability of a disruption in an appointment. These methods utilize past behaviors of individuals for the estimation of future no-show and cancellation probability. Time series methods forecast future events such as no-shows and cancellations based on the past events by using stochastic models. There are different types of time series models; the common three classes amongst them are: the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. These three classes depend linearly on previous data [Brockwell 2009]. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. Smoothing is an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures and rapid phenomena. Many different algorithms are used in smoothing. Some of the most common algorithms are the moving average, and local regression [Simonoff 1996]. Bayesian inference is a method of statistical inference in which some kind of evidence or observations are used to update its previously calculated probability such as improving the initial estimate of disturbances probabilities [Bolstad 2007]. To use Bayes' theorem, we need a prior distribution $g(p)$ that gives our belief about the possible values of the parameter p before incorporating the data. The posterior distribution is proportional to prior distribution times likelihood $f(y|p)$:

$$g(p \mid y) \propto g(p) \times f(y \mid p) \tag{1-1}$$

If the prior is continuous, the posterior distribution can be calculated as follows:

$$g(p|y) = \frac{g(p) \times f(y|p)}{\int_0^1 g(p) \times f(y|p) dp} \tag{1-2}$$

While individual based methods are fast and effective in modeling the behavioral (no-show) pattern of each individual, and work well with a small dataset, they do not use the predictive information from the rest of the population and hence do not provide a reliable initial estimate of no-show and cancellation probabilities which is especially important in our problem. In chapter 2, the performance of some of above methods will be compared with the proposed work.

As described above, each of population based and individual based approaches have some advantages and disadvantages. However, none of the studies in the literature have considered using these methods together in order to overcome their problems and improve their performance. In chapter 2, a hybrid probabilistic model will be developed that combines logistic regression as a population based approach along with Bayesian inference as individual based approach for no-show and cancellation prediction. To demonstrate its effectiveness, the proposed model will be compared to the representative algorithms from both population based and individual based approaches

1.2. Preliminaries

This Section introduces some of the preliminaries required to comprehend the proposed algorithm. First the notations used in the study are described. Next, some basics about logistic regression, Beta and Dirichlet distributions as the vital components of the proposed model are explained. Finally, more details about Bayesian update of Beta and Dirichlet distributions as the main procedure of the proposed algorithm for modeling the individual's behavior are provided.

1.2.1. Notations Used for the Probability of Disturbance Prediction Model

Table 1 describes the notations used for the proposed prediction model.

Table 1: Notations used for the proposed prediction model

Notations	Description
i,j,k	Indices of individual $(\mathrm{i} = 1, \dots \mathrm{n})$, appointment no. j $(\mathrm{j} = 1, \dots, \mathrm{J})$, and attendance record of type k $(\mathrm{k} = 0, \dots, \mathrm{K})$
D_{GI}	Database of each individual's personal information
D_{NR}	Database of appointment information and attendance records of each person
$F(X_i, B_k)$	Logistic regression model
B_k	Vector of logistic regression parameters for attendance record of type k $(B_k = [\beta_{k0}, \beta_{k1}, \dots, \beta_{kl}])$
X_{ij}	Factors affecting probability of attendance of person i for appointment j (independent variables in the logistic regression model),$\left(X_{ij} \in D_{GI} \vee D_{NR}, X_{ij} = \left[x_{ij0}, x_{ij1}, .., x_{ijl},\right]\right)$
Y_{ij}	Person i attendance type for appointment j (No-show, cancellation, show-up)
$\hat{p}_{ik}^{0}$	Initial estimate for probability of attendance of type k for person i
$\left(\alpha_{ij}^{pos}, \beta_{ij}^{pos}\right)$	Beta distribution posterior parameters of person i probability of attendance for appointment j
$\hat{p}^{Model}$	Estimated probability of attendance by the model
$\hat{p}^{Emp}$	Real (empirical) probability of attendance
W_j	Weight for Appointment j
T	Threshold for convergence of the objective function
D	Improvement in the objective function at each iteration $\left(\Delta\, p - value_{paird\ F-test}\right)$
LG	Logistic regression model

1.2.2. Binomial and Multinomial Logistic Regression

Logistic regression is a generalized linear model used for binomial regression, which predicts the probability of occurrence of an event by fitting numerical or categorical predictor variables in data to a logit function [Agresti 2002]:

$$logit(p) = log(p/1 - p) \tag{1-3}$$

where $0 \leq p \leq 1$ and $(p/1 - p)$ is the corresponding odds. The logistic function can be written as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}} \tag{1-4}$$

where p represents the probability of a particular outcome. Given the set of explanatory variables and unknown regression coefficients $\beta_j, (0 < j < k)$ can be estimated using maximum likelihood (MLE) methods common to all generalized linear models [Hilbe 2009].

Multinomial logistic regression is a generalization of the binomial model used when the dependent variable follows a multinomial distribution. The model then takes the form:

$$\begin{cases} P_k = \frac{exp(XB_k)}{1+\sum_{k=0}^{K} exp(XB_k)}, k = 1,2, \ldots, K \\ P_0 = \frac{1}{1+\sum_{k=0}^{K} exp(XB_k)} \end{cases} \tag{1-5}$$

where P_k is the probability of k^{th} event and X is the vector of explanatory variables. The unknown vector of parameters B_k is typically estimated by the maximum a posteriori (MAP) estimation, which is an extension of maximum likelihood using regularization of the weights [Agresti 2002].

1.2.3. Beta and Dirichlet Distributions

Beta distribution: $Beta\,(\alpha,\beta)$ represents a family of common continuous distributions defined on the interval [0,1] parameterized by two positive shape parameters, typically denoted by α and β with probability density function:

$$f(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{1-6}$$

where Γ is the gamma function, and $\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$ is a normalization constant to ensure that the total probability integrates to unity. The beta distribution is the conjugate prior of the binomial distribution. From the Bayesian statistics viewpoint, a Beta distribution can be seen as the posterior distribution of the parameter p of a binomial distribution after observing $\alpha-1$ independent events with probability p and $\beta-1$ with probability $1-p$, if there is no other information regarding the distribution of p [Evans et al. 2000].

Dirichlet distribution (denoted by $Dir(\alpha)$) is the generalization of beta distribution to a family of continuous multivariate probability distributions parameterized by the vector α of positive reals. The Dirichlet distribution of order $K \geq 2$ with parameters $\alpha_1,\ldots,\ \alpha_K > 0$ has a probability density function with respect to Lebesgue measure on the Euclidean space R^{K-1} given by [Evans et al. 2000]:

$$f(\gamma_1,..,\gamma_K,\alpha_1,\ldots,\alpha_K)=\frac{1}{B(\alpha)}\prod_{k=1}^{K}\gamma_k^{\alpha_k-1} \tag{1-7}$$

for all $\gamma_1,\ldots,\gamma_K > 0$ satisfying $\sum_{i=1}^{K}\gamma_i < 1$ (our work incorporates $K=3$ which is based on the number categories: no-show, cancellation, and show-up). The density is zero outside this open $(K-1)$-dimensional simplex. The normalizing constant is the multinomial beta function, which can be expressed in terms of the gamma function:

$$B(\alpha) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K} \alpha_k)}, \alpha = (\alpha_1, \dots, \alpha_K) \tag{1-8}$$

Dirichlet distribution is the multivariate generalization of the beta distribution (multinomial distribution), and conjugate prior of the categorical distribution and multinomial distribution in Bayesian statistics. That is, its probability density function returns the belief that the probabilities of K rival events given that each event has been observed $\alpha_j - 1$ times.

1.2.4. Bayesian Update of Beta and Dirichlet Distributions

In Bayesian statistics, a Beta distribution [Bolstad 2007] is a common choice for updating a prior estimate of the Binomial distribution parameter p because:

1. A Beta distribution is the conjugate prior of a Binomial distribution (See Section 1.2.3).
2. Unlike a Binomial distribution, a Beta distribution is a continuous distribution, which is much easier to work with in terms of inference and updating.
3. A Beta distribution has two parameters, which allows it to take different shapes, making it suitable for representing different types of priors.

If $Beta(\alpha, \beta)$ is used as a prior, based on the conjugacy property of Beta distribution, the posterior would be a new Beta posterior with parameters $\alpha' = \alpha + y$ and $\beta' = \beta + n - y$. In other words, Beta distribution can be updated simply by adding the number of successes y to α and the number of failures $n - y$ to β:

$$g(p \mid y) \sim Beta\left(\alpha + y, \chi + n - y\right) \tag{1-9}$$

$$g(p \mid y) = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(y + \alpha)\Gamma(n - y + \beta)} p^{y - \alpha - 1}(1 - p)^{n - y + \beta - 1} \tag{1-10}$$

As discussed earlier, individual-based approaches like empirical Bayesian inference will not be able to provide an initial estimate of the prior distribution. Hence, before applying the Bayesian update, the parameters of the prior distribution should be initialized.

[Bolstad 2007] suggests choosing parameters that match the belief about the location (mean) and scale (standard deviation) of the original distribution. Hence, if an initial guess of parameter p is available, which in our study can be obtained from population-based approaches such as logistic regression, Beta distribution prior parameters can be computed by solving the following system of Equations for α and β.

$$\begin{cases} p_i = \dfrac{\alpha}{\alpha+\beta} \\ \dfrac{p_i(1-p_i)}{n} = \sqrt{\dfrac{p_i(1-p_i)}{\alpha+\beta+1}} \end{cases} \tag{1-11}$$

The point estimate of the posterior parameter p of the binomial distribution would be the mean of Beta distribution $\dfrac{\alpha}{\alpha+\beta}$ of the updated Beta distribution.

Similarly, Dirichlet distribution is a regular option for updating prior estimate of Multinomial distribution parameters $\alpha = (\alpha_1, \dots, \alpha_K)$. To use Bayes' theorem, we need a prior distribution $g(\alpha^{pri})$ that gives our belief about the possible values of the parameter vector $\alpha = (\alpha_1, \dots, \alpha_K)$ before incorporating the data.

Based on earlier discussion on *Dir(α)* , $\alpha = (\alpha_1, \dots, \alpha_K)$ can be used as prior density, which results in a new Dirichlet posterior with parameters vector $a_k^{pos} = a_k^{pri} + y_k$, where y_k is the number of occurrences of each category in the incorporated data. In other words, Dirichlet distribution can be updated simply by adding the new occurrence number of each category to the prior parameters α_k:

$$g(\alpha|\gamma) = \frac{\Gamma(\sum_{k=1}^{K} y_k+\alpha_k)}{\prod_{k=1}^{K} \Gamma(y_k+\alpha_k)} \prod_{k=1}^{K} \gamma_{\mathrm{k}}^{\alpha_k+y_k-1} \tag{1-12}$$

The posterior mean would then be $E(a_k|y_1, \dots, y_K) = \frac{y_k+a_k}{\sum_{k=1}^{K} y_k+\sum_{k=1}^{K} a_k}$ with variance:

$$Var(\alpha_k|y_1, \dots, y_K) = \frac{(y_K+\alpha_k)\left((\sum_{k=1}^{K} y_k+\sum_{k=1}^{K} \alpha_k)-(y_K+\alpha_k)\right)}{\left((\sum_{k=1}^{K} y_k+\sum_{k=1}^{K} \alpha_k)^2 (\sum_{k=1}^{K} y_k+\sum_{k=1}^{K} \alpha_k+1)\right)} \tag{1-13}$$

For choosing an extended version of the procedure used for Beta distribution can be applied by letting $\alpha_k^{pri} = P_k$; where P_k is the output of the multinomial logistic regression. As an alternative to above procedure, several researchers [Leonard 1973; Aitchison 1985; Goutis 1993; Forster and Skene 1994] proposed using a multivariate normal prior distribution for multinomial logits.

1.2.5. Hotelling's T^2 Distribution

In statistics, Hotelling's T^2 statistic [Evans et al. 2000] is a generalization of Student's *t*-statistic that is used in multivariate hypothesis testing. Hotelling's T^2 statistic is defined as follows:

$$t^2 = n(\rho - \mu_p)^T W^{-1}(\rho - \mu_p) \tag{1-14}$$

where *n* is a number of points (see below), ρ is a column vector of K elements and W is a $p \times p$ sample covariance matrix. If $\rho \sim N_K(\mu, V)$ is a random variable with a multivariate Gaussian distribution and $W \sim W_K(V, n-1)$ (independent of ρ) has a Wishart distribution with the same non-singular variance matrix V and $n - 1$, then the distribution of t^2 is, Hotelling's T^2 with parameters K and *n,* where F representing F-distribution:

$$\frac{n-K}{n(n-1)} t^2 \sim F_{K,n-K} \tag{1-15}$$

T^2distribution can be used for pairwise comparison of the mean of two sets of multidimensional data H_0: $\mu_D = 0$ against H_1: $\mu_D \neq 0$, e.g. estimated and empirical probabilities of patient attendance. Above hypothesis can be tested using F-statistic $\frac{n-K}{n(n-1)} t^2 \sim F_{p,n-p}$ where $t^2 = n\bar{P}_D' S_{P_D}{}^{-1} \bar{P}_D$ where and S_{P_D} is calculated as follows:

$$S_{P_D} = \frac{1}{n-1} \sum_{i=1}^{n} \left(P_{D_i} - \bar{P}_D\right)\left(P_{D_i} - \bar{P}_D\right)' \tag{1-16}$$

Also $\bar{P}_D = \frac{1}{n} \sum_{i=1}^{n} P_{Di}$ where $P_{Di} \leftarrow \left[p_{i1}^{Emp} - \hat{p}_{i1}^{Model}, \dots, p_{iK}^{Emp} - \hat{p}_{iK}^{Model}\right]$ is the vector of pairwise differences between person i empirical and estimated probabilities for type $k \in (1, \dots, K)$ attendance, e.g. probabilities of no-show, cancellation and show-up.

CHAPTER 2 PREDICTING DISTURBANCES IN APPOINTMENT SCHEDULING THROUGH HYBRID PROBABILISTIC MODELING

2.1. The Proposed Algorithm for No-Show Prediction

Algorithm 1 illustrates the flow of the steps taken by the proposed approach for estimating no-show probability which can be categorized in three stages:

1. Initial no-show probability estimation
2. Bayesian update of the no-show estimate
3. Weight optimization

Algorithm 1: No-show Prediction Algorithm

Input: Input data $\left(X_{ij}, Y_{ij}\right)$, Threshold parameter T

Output: Estimated no-show probability $\hat{p}^{Model}$, Beta distribution posterior parameters$\left(\alpha_{ij}^{pos}, \beta_{ij}^{pos}\right)$, Logistic regression estimated parameters $\hat{B}$

Procedure:

```
/* Logistic regression*/
$\hat{B} \leftarrow$ Calculate MLE of Equation (1.4) parameters
$\hat{p}_{0ij}\left(Y_{ij} = 1 | X_{ij}\right) \leftarrow F\left(X_{ij}, \hat{B}\right)$
$\left(\alpha_i^{pri}, \beta_i^{pri}\right) \leftarrow$Solve system of Equation (1-11) with $\hat{p}_{0i}(Y_i = 1 | X_i)$
/*Weight optimization*/
$\hat{p}_i^{Real} \leftarrow \frac{\sum_{j=l}^{m} Y_{ij}}{m - l + 1}$
Until Equation (2-1) improvement $D < T$ do
    $W_j \leftarrow$set a value for appointments weights
    /*Bayesian update */
    $\begin{cases} \alpha_{ij}^{pos} \leftarrow \alpha_i^{pri} + \sum_{i,1}^{i,j-1} \left( \prod_{\omega \in W} w_{ij\omega} \right) Y_{ij} \\ \beta_{ij}^{pos} \leftarrow \beta_i^{pri} + n_i - \sum_{i,1}^{i,j-1} \left( \prod_{k \in W} w_{ijk} \right) Y_{ij} \end{cases}$
    $\hat{p}^{Model} \leftarrow \frac{\alpha_{ij}^{pos}}{\alpha_{ij}^{pos} + \beta_{ij}^{pos}}$
    $\bar{P} \leftarrow \sum_{i=1}^{n} \left(\hat{p}_i^{Model} - \hat{p}_i^{Real}\right) / n$
    $S_P \leftarrow \frac{\sum_{i=1}^{n} \left(\hat{p}_i^{Model} - \hat{p}_i^{Real}\right)^2 - \left[\left(\sum_{i=1}^{n} \left(\hat{p}_i^{Model} - \hat{p}_i^{Real}\right)\right)^2 / n\right]}{n - 1}$
    $t_0 \leftarrow \bar{P} / S_P / \sqrt{n}$
Return $\hat{p}^{Model}$
```

In the first stage, based on the dataset of individuals' personal information (D_{GI}), (such as gender, marital status, etc.) and their sequence of appointment information (e.g. previous attendance records (D_{NR})), a logistic regression model $F\left(X_{ij}, \hat{B}\right)$ is formulated (*line 2*). Then, using logistic regression, an initial estimate of no-show probability is calculated, given by $\hat{\mathrm{p}}_{0\mathrm{i}}(\mathrm{Y}_{\mathrm{i}} = 1|\mathrm{X}_{\mathrm{i}})$. As discussed in Section 1.1.2, Logistic regression bundles the information of the complete population together and finds a reliable initial estimate of no-show $(\hat{p}_{0i})$.

In the second stage, which is interlaced with the third stage, the initial estimate is used in a Bayesian update procedure to find the posterior no-show probability for each person. For this purpose, $\hat{p}_{0i}$ is transformed into prior parameters of a Beta distribution $\left(\alpha_i^{pri}, \beta_i^{pri}\right)$ as shown in *line 4.* Next, using the attendance record of each person $\left(Y_{ij}\right)$ the posterior parameters $\left(\alpha_i^{pos}, \beta_i^{pos}\right)$ and posterior probability of no show $\hat{p}^{Model}$ is calculated (*lines 10 and 11*). As discussed in Section 1.1.3, the reason Bayesian update procedure is applied to the output of logistic regression is that, typically regression models cannot consider individual patients behavior. Also, updating regression parameters based on new data records is both difficult and only marginally effective (especially when the model is already constructed on a huge dataset) in comparison to Bayesian update.

In the third stage, appointments are weighted based on on a subset of factors $W= [w_1, \dots, w_\omega]$ *(line 8)* to increase the model performance in estimating the real probability of no-show. An optimization procedure is used for finding the optimal value of the weights. The objective function of the model is to minimize the difference between the real and model estimated probability of no-show:

$$\begin{aligned} &\max p - value_{paired-t-test} = p\left(t_0 < -t_{\frac{\alpha}{2},n-1}\right) + p\left(t_0 > t_{\frac{\alpha}{2},n-1}\right) \\ &S.T: \\ &w_1,\dots,w_\omega \in (0,1) \end{aligned} \tag{2-1}$$

Where $w_1, \dots, w_\omega$ are the weights to be optimized and $p - \text{value}_{paird\ t-test}$ is the p-value of a two sided statistical hypothesis testing of the paired estimated p using the model and estimated p using the attendance records:

$$\begin{cases} H_0 : p_D^{Model} = p_D^{\text{Re}\,al\ Datal} \\ H_1 : p_D^{Model} \neq p_D^{\text{Re}\,al\ Datal} \end{cases} \tag{2-2}$$

It should be noted that the mean squared error (MSE) can also be used as the objective function. However, t-statistics which is used above not only contains MSE in itself (S_P in the denominator of t statistics is a linear function of MSE) (*line 14*), but also has a statistical distribution which makes it better choice for our optimization model.

In (2-1), $t_{\frac{\alpha}{2},n-1}$ is the percentage of points or value of *t* random variables with *n-1* degrees of freedom such that the probability that t_{n-1} exceeds this value is α, and $t_0 = \frac{\bar{P}}{S_P/\sqrt{n}}$ where $\bar{P} = \sum_{i=1}^{n}(\hat{p}_i^{Model} - \hat{p}_i^{Real})/n$ and S_P is calculated as follows:

$$S_p = \frac{\sum_{i=1}^{n}\left(\hat{p}_i^{Model} - \hat{p}_i^{\text{Re}\,al}\right)^2 - \left[\left(\sum_{i=1}^{n}\left(\hat{p}_i^{Model} - \hat{p}_i^{\text{Re}\,al}\right)\right)^2 \Big/ n\right]}{n-1} \tag{2-3}$$

Where $\hat{p}_i^{Real}$ is the real rate of no-show for person i calculated as $\hat{p}_i^{Real} \leftarrow \frac{\sum_{j=l}^{m} Y_{ij}}{m-l+1}$, with Y_{ij} as a binary (random) variable representing records of no-show/show of patient i for appointment j. Here, l is the index of first appointment in the validation dataset which is discussed shortly, and m is the total number of appointments in the validation dataset for patient i. Also $\hat{p}_i^{Model}$ is the estimated no-show probability calculated based on weighted appointments using the proposed model.

The optimization procedure is as follows: at every iteration, a vector of weights is assigned to the appointments in validation dataset *(line 8)*. The weighted appointments are then plugged into the Bayesian update mechanism for estimating the probability of no-show *(lines 10 and 11)*. Next, the estimates of the proposed model and real attendance records are compared by forming a t- statistic *(lines 12 to 14)* and the p-value of the paired t-test which shows the goodness of the assigned weights, is used for improving the initial set of weights *(line 7)* . This procedure continues until no improvement is observed. Then, the $\hat{p}^{Model}$ of the iteration resulted in the best value of objective function is used as the no show estimate.

2.2. Experimental Results

Here the proposed method is evaluated based on a healthcare dataset of 99 patients at the Veteran Affairs (VA) Medical Center in Detroit. The dataset includes the following data from patients' personal and appointment information: (1) sex, (2) date of birth (DOB), (3) marriage status, (4) medical service coverage, (4) address (zip code), (5) clinic and (6) attendance record.

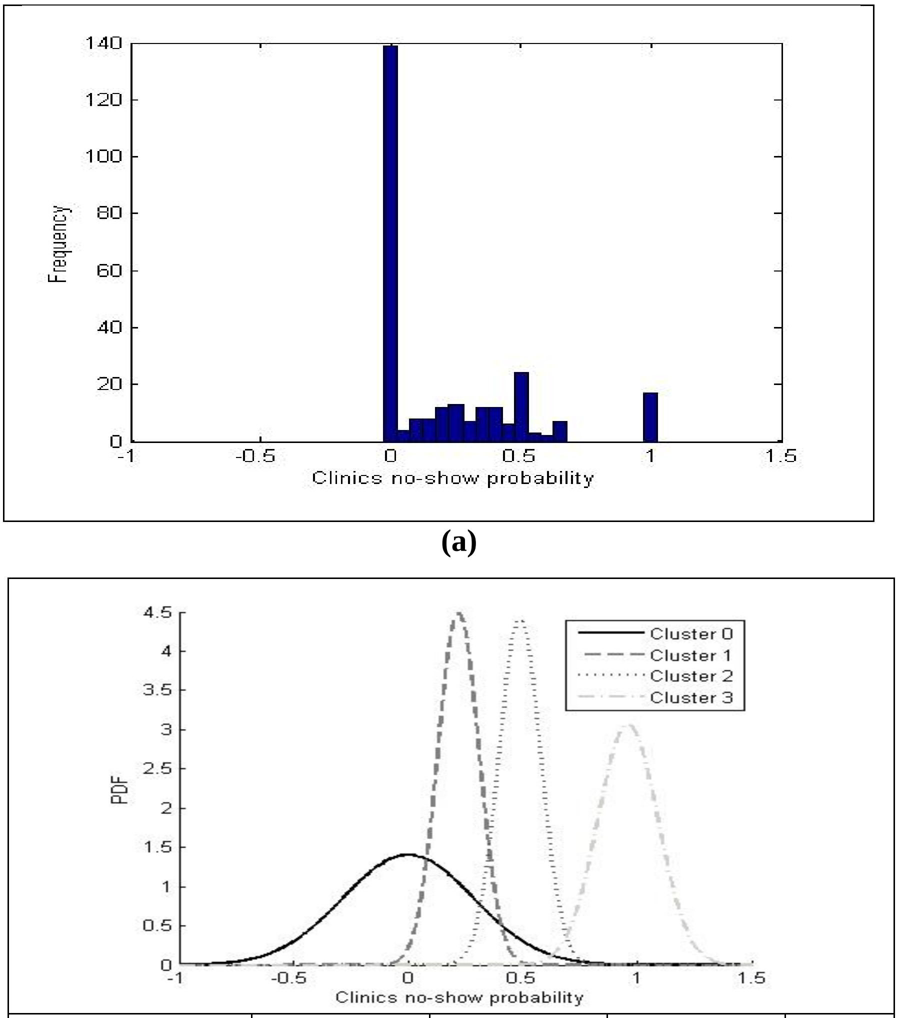
This Section is organized as follows: first data processing is discussed. Next, a stylized example for one patient is presented to show how the model works. Finally, the results of applying the model to the dataset is discussed using two types of analysis: one by defining training, validation and test dataset on patients, and one on appointment time.

2.2.1. Data Preprocessing

The data attributes in the dataset should be preprocessed before being used in the model. This includes: coding, dealing with missing attributes, and co-linearity elimination. Besides, because of the variety of clinics (more than 150 in our case), if this explanatory variable gets directly used in the model, the accuracy of the logistic regression will be severely affected.

Such problem can be addressed by clustering similar clinics respect to their no-show rate. While various types of clustering algorithms can be used for this purpose, since the clinics are originally different in type, grouping them into a set of clusters will result in clusters with different density and dispersion. Such characteristics can be effectively considered using Generalized Mixture Models (GMM) [Alpaydim 2010].

Figure 1(a) illustrates the histogram of clinics' no-show probability and Figure 1 (b) shows the result of clustering the clinics based on their no-show probability using GMM. The final result which is four clusters has been verified by a team of experts.



Clinic Cluster	0	1	2	3
μ	**0**	**0.2197**	**0.4874**	**0.9595**
σ	**0.2843**	**0.0890**	**0.0904**	**0.1300**

(b)

Figure 1: (a) The histogram of clinics no-show probability, (b) The result of clustering the clinics based on their no-show probabilities

Also, (1) appointment recency, (2) appointment closeness to non-working days (Saturday, Sunday, and holidays), and (3) clinic cluster, are considered as weighting factors ($W = [w_1, w_2, w_3]$). Regarding the first factor, it is reasonable that no-show records that occurred long time ago do not carry the same weight as recent no-shows. This is based on the fact that patients may gradually or abruptly change their behavior, which should be reflected in the model.

Regarding the second and third weighting factors, the study of data revealed a strong correlation among no-show rate and days close to holidays and clinic clusters.

The weights discussed above are arranged in a special data structure before being applied to the data. For the appointment recency where more importance should be assigned to the recent appointments a logarithmic time framework with five weights is considered. For the appointment closeness to non-work days two weights are applied: one for Monday to Thursday and one for Friday and days before holidays. Finally, for clinics cluster, based on the groups derived using GMM, four weights are defined. Table 2 shows the final data structure and optimal value of the weights which is gained by solving Equation (2-1) using Genetic Algorithm (GA) algorithm.

Table 2: Data structure for the weighting factors

Appointment recency					**closeness to non-work days**		**Clinic cluster**			
<1 wk.	<1 mon.	<3 mon.	6< mon.	> 6 mon.	Not-before holiday	Before holiday	Very important	…	…	Not important
1	1	1	0.95	0.9	1	.925	1	…	…	0.75.

2.2.2. Applying the Proposed Model to a Sample Patient

Here the procedure of no-show probability estimation for a randomly selected patient is explained. The selected patient is male, born on 02/5/1978, never married, degree of medical service coverage less than 5% with zip code 48235. Table 3 shows his appointment information as patterns of show/no-show from 10/13/2009 to 12/31/2009 (training dataset). Note that no-shows are represented by 1 while shows are represented by 0.

Table 3: Attendance record of a sample patient

Appointment No.	Appointment date	Clinic cluster	No-show
1	10/13/2009	1	**0**
2	10/29/2009	1	**1**
3	11/10/2009	0	**0**
4	11/17/2009	1	**1**
5	12/2/2009	2	**1**
6	12/8/2009	1	**0**
7	12/9/2009	2	**0**
8	12/23/2009	1	**0**
9	12/23/2009	1	**1**
10	12/29/2009	1	**0**
11	12/31/2009	0	**1**

Using patient personal and appointment information as well as the attendance record, the parameters of the fitted logistic regression model are calculated in Table 4.

Table 4: A sample logistic regression model fitted to the dataset

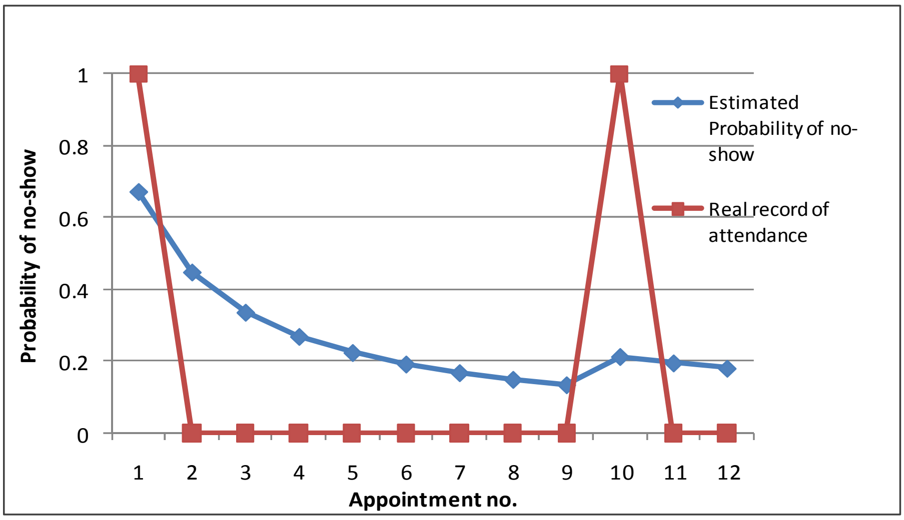
Sex	DOB	Marriage status	Medical service coverage	Zip code	Clinic cluster	Recency	Closeness to non-workday	Constant
71.6917	-0.8600	6.51E-05	-0.13596	0.0180	0.0015	0.4822	0	3.0410

Based on the estimated coefficients of logistic regression the probability of not showing up in the first appointment in the testing dataset (1/25/2010) is estimated as p=0.3453. This estimate is used for building the prior *Beta*(0.3453,0.6547) of the Bayesian updating procedure by solving Equation (1-11). Table 5 illustrates the updated parameters of Beta distribution as well as the estimated probability of no-show after each appointment.

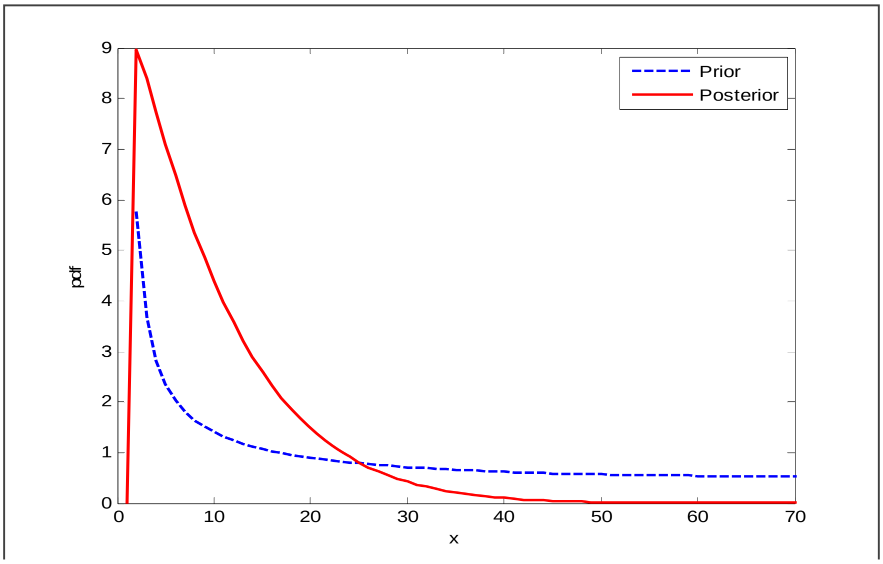
Table 5: Bayesian update of Beta distribution parameters

Appointment No.	Appointment date	Clinic Category	Weight			No-show	Weighted No-show	α	β	Estimated p
			Closeness to non-workday	Recency	Clinic cluster					
								0.3453	0.6547	**0.345**
12	1/25/2010	0	1	0.9	1	1	0.35	0.695	0.655	**0.515**
13	1/26/2010	1	1	0.9	0.9	0	0	0.695	1.655	**0.296**
14	2/2/2010	0	1	0.9	1	0	0	0.695	2.655	**0.208**
15	2/4/2010	2	1	0.9	0.75	0	0	0.695	3.655	**0.160**
16	2/6/2010	2	1	0.9	0.75	0	0	0.695	4.655	**0.130**
17	2/17/2010	0	1	0.9	1	0	0	0.695	5.655	**0.109**
18	2/18/2010	1	1	0.9	0.9	0	0	0.695	6.655	**0.095**
19	2/23/2010	0	1	0.9	1	0	0	0.695	7.655	**0.083**
20	3/2/2010	1	1	0.9	0.9	0	0	0.695	8.655	**0.074**
21	3/9/2010	0	1	0.9	1	1	0.35	1.045	8.655	**0.108**
22	3/16/2010	0	1	0.9	1	0	0	1.045	9.655	**0.098**
23	3/18/2010	2	1	0.9	0.75	0	0	1.045	10.655	**0.089**

As graphically illustrated in Figure 2 (a), the Bayesian update reacts quickly to each new data record, which means that the procedure can rapidly converge to the real distribution of no-show. Figure 2 (b) compares the prior and posterior distributions of no-show probability before and after applying testing data. It is easy to follow how the mass of the density function has moved to the left in the posterior which can be interpreted as decreasing the probability of no-show.



(a)



(b)

Figure 2: Applying the proposed model for a sample patient :(a) Real record of attendance and estimated probability of no-show using the proposed model (b) Prior and posterior of Beta distribution for modeling no-show

2.2.2.1. Time wise Analysis

In this Section, the performance of the proposed model is compared with a number of population and individual based methods based on time wise analysis. In this regard the training, validation and testing data are defined as follows: appointments occurred before 6/31/2009 for training, appointments between 6/31/2009 and 12/31/2009 for validation, and appointments after 12/31/2009 for testing. Such a setting is used for all of the time-wise experiments.

The comparing methods including: Box smoothing, autoregressive integrated moving average model (ARIMA), decision tree, and multiple logistic regression with same predictors as used in the proposed model regression part and rule-based methods. The moving window size of Box method is checked for the range of 1 to 7, where only 5 is considered. For ARIMA model two cases ARMA (1,1,0) and ARMA(2,2,0) are considered. Also, J48 and PART algorithms are used for building the decision tree rule-based methods.

Figure 3 compares the Mean Squared Error (MSE) of the comparing methods. Based on MSE measure the proposed model performs clearly better than other methods while rule-based has the worst result. As can be seen from the results, in general, individual-based methods outperform population based methods, while bundling these methods together (as we have done in the proposed methods) significantly better method than both.

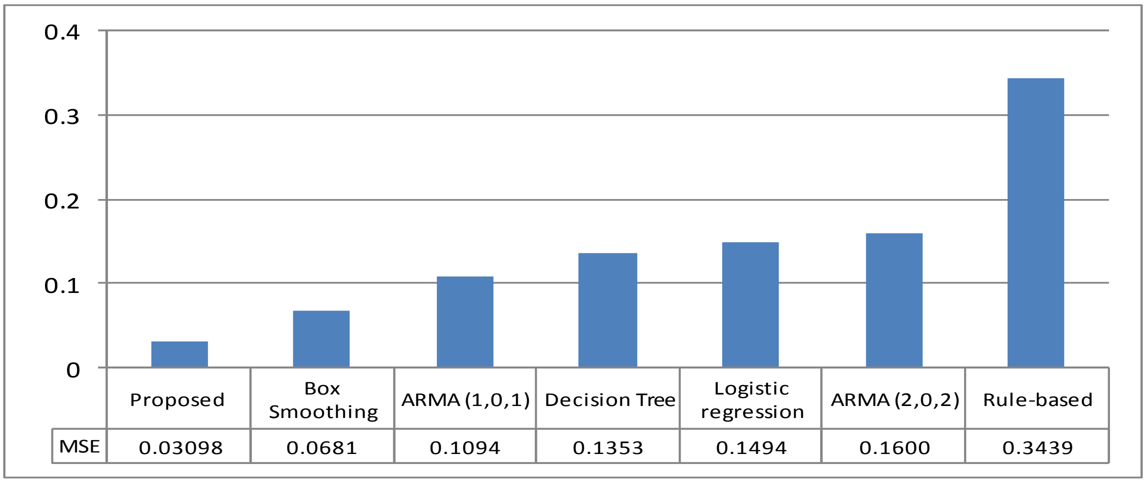
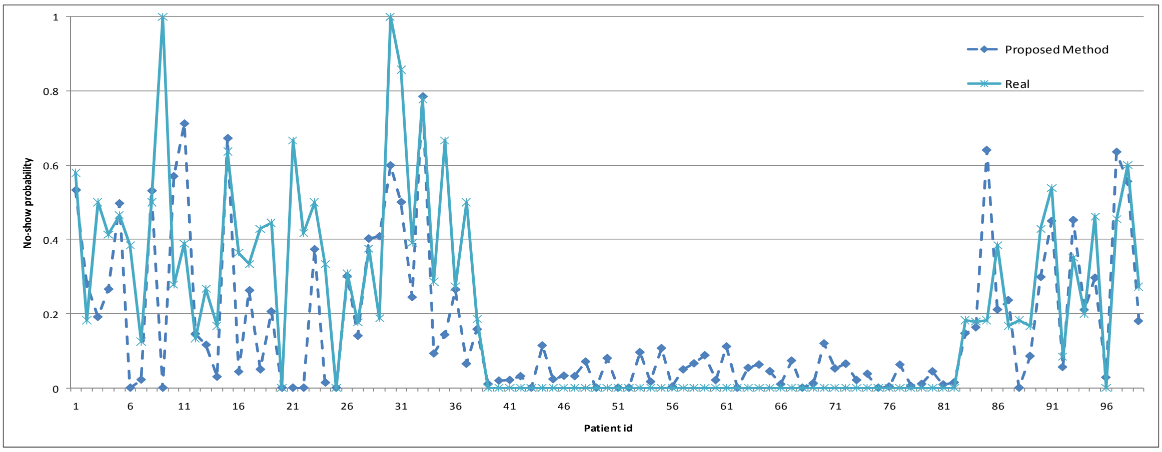
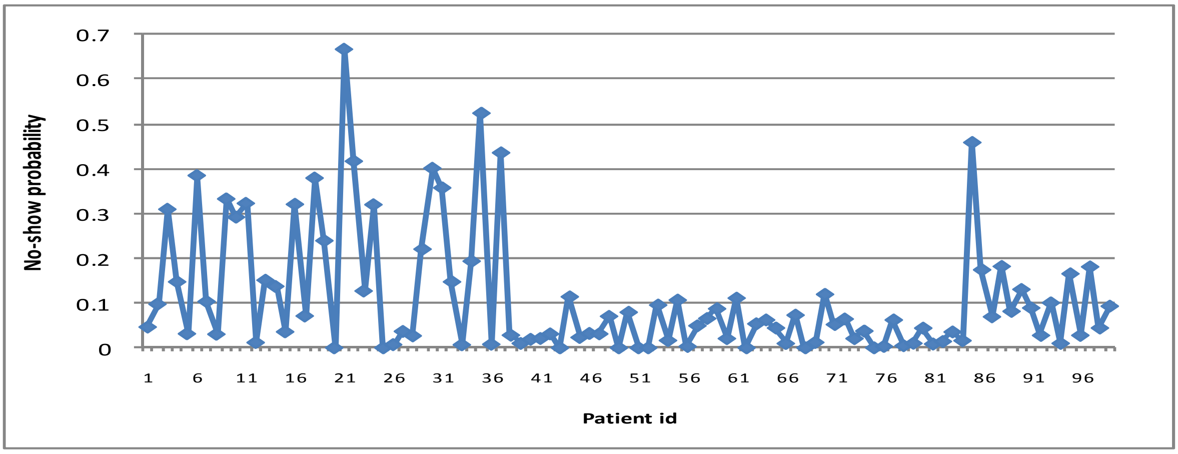


Figure 3: Mean Squared Error (MSE) of the other methods used for comparison

Figures 4 to 6 illustrate some of the comparing methods estimates versus real probability of no-show over different patients. As can be seen from Figure 4 (a), the proposed approach estimates often follows the real pattern correctly. This is better illustrated in Figure 4 (b) which shows the absolute difference between the estimated and real probability of no-show. Here, the mean of differences is 0.1104 which is acceptably low. There are also few cases with absolute difference more than 0.5. Later analysis reveals fewer available data records for those patients.



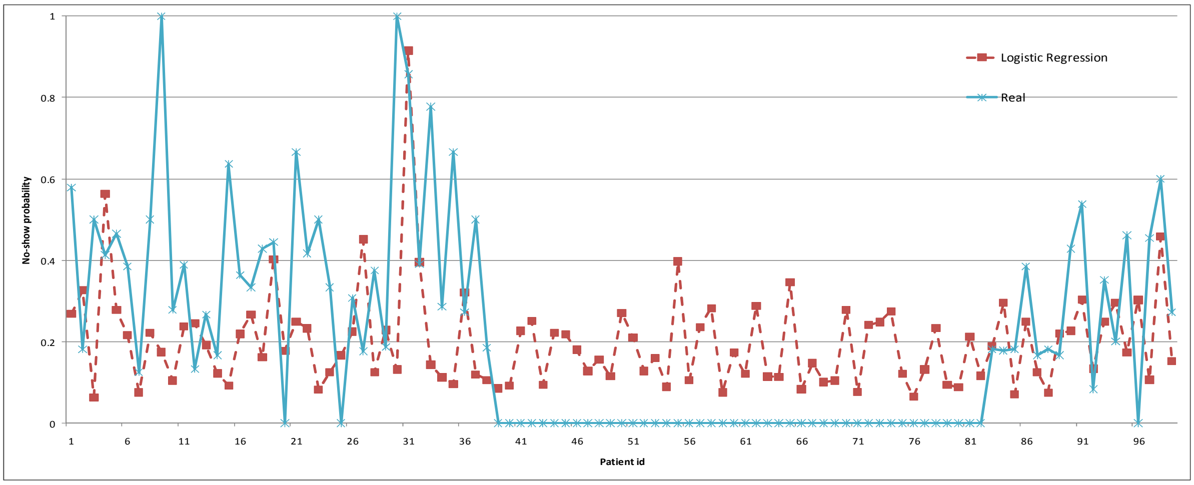
(a)



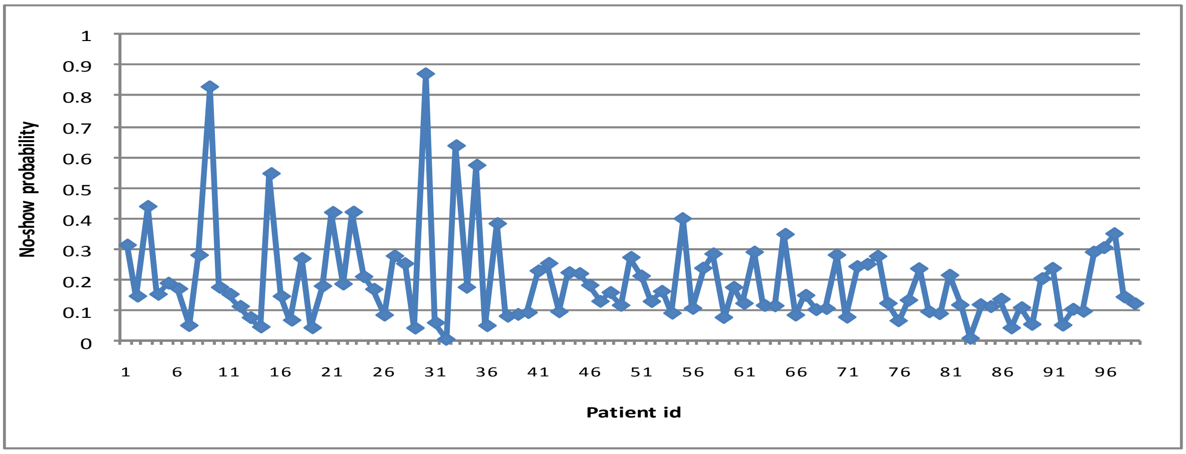
(b)

Figure 4: Proposed approach performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability

Figure 5 (a) illustrates the estimates from one of the population based methods which is logistic regression. As can be seen, the estimates tend to have small fluctuations around an approximately fixed mean. Such result clearly shows that the regression models may not fully capture the difference among patients' personal behaviors. The absolute difference between the estimated and real probability of no-show which is shown in Figure 5 (b) also confirms similar results. Here the mean of the differences is 0.1935 while the maximum difference is 0.8683 which is considerable. Such a result is similar to other population based methods discussed earlier.



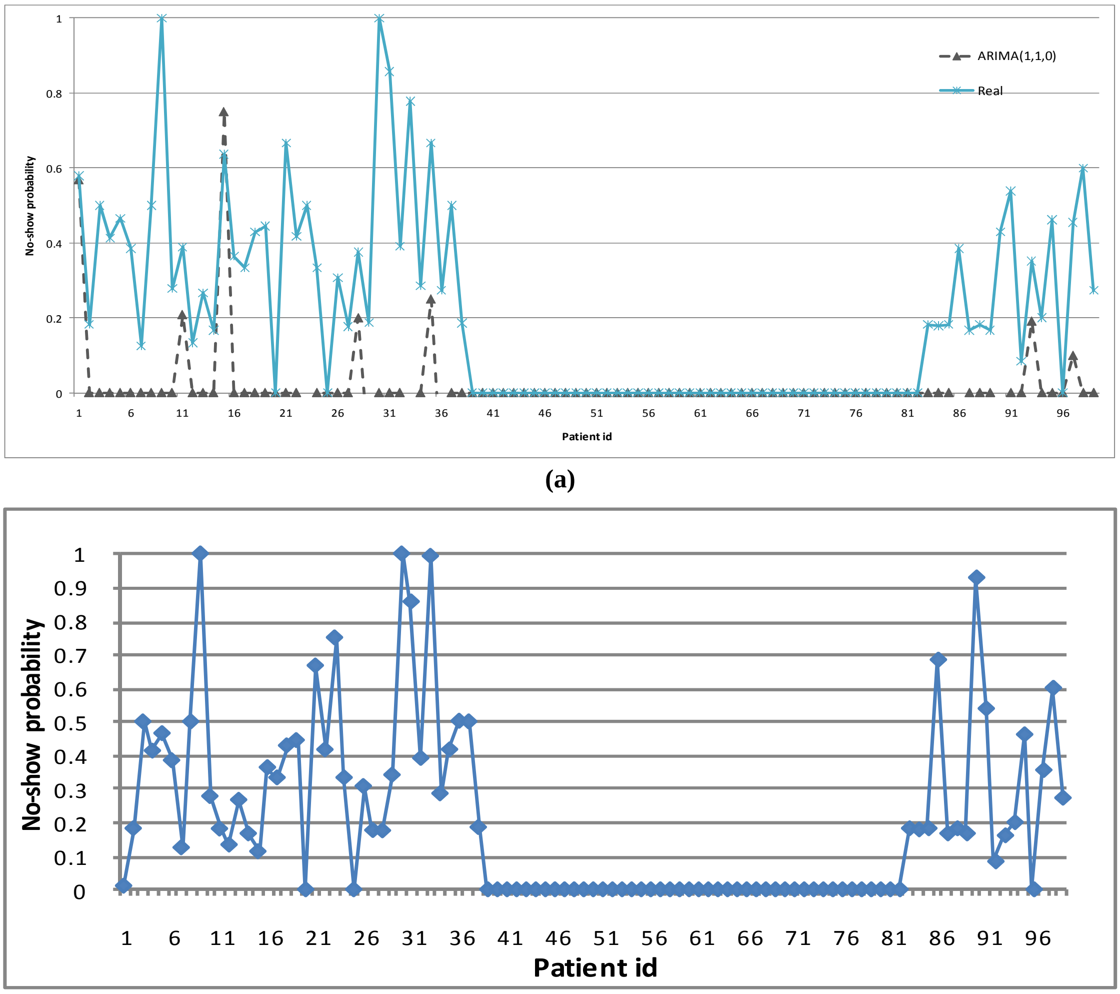
(a)



(b)

Figure 5: Logistic regression performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability

Finally, Figure 6(a) shows the results from ARIMA (1,1,0) model which is one of the individual based methods. As can be seen for a large portion of patients that have real no-show rate of larger than zero ARIMA can barely follow the real pattern. This can also be checked in Figure 6 (b) which has several differences greater than 0.5 and a few differences equal to 1.



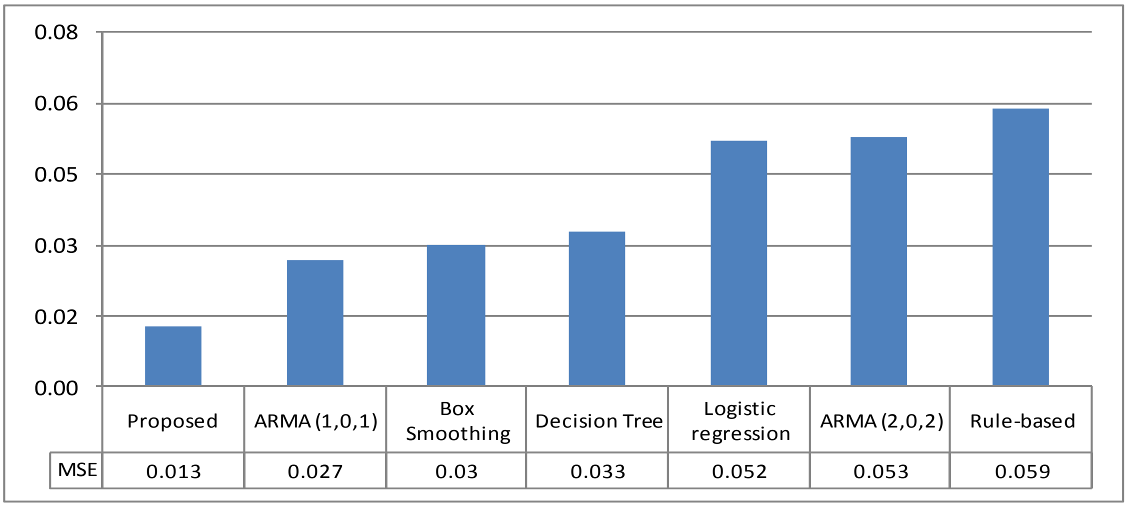
(a)

(b)

Figure 6: ARIMA (1,1,0) performance over patients: (a) estimated versus real probability of no-show, (b) Absolute difference of estimated and real no-show probability

2.2.2.2. Patient wise Analysis

Here the comparing methods discussed in previous Sections are studied based on a patient wise analysis. For this purpose, out of 99, 50 patients are randomly chosen and used for training, 20 are randomly selected for validation and the 29 are used for testing. Figure 7 illustrate the results which is similar to time wise analysis of the comparing methods.



	Proposed	ARMA (1,0,1)	Box Smoothing	Decision Tree	Logistic regression	ARMA (2,0,2)	Rule-based
MSE	0.013	0.027	0.03	0.033	0.052	0.053	0.059

Figure 7: Mean squared error (MSE) of different methods used for comparison comparing methods

The results from Figures 2 to 7 clearly show the capability of the proposed model in estimating probability of non-attendance for both current and new patients of a health care system.

2.2.2.3. Discussion

In this Section, up to this point, a probabilistic model based on logistic regression and Bayesian inference has been developed to estimate the patients' no-show probability in real-time. Also, the effects of appointment date and clinic on the proposed method have been modeled. Next, based on a dataset from a Veteran Affair medical center, the effectiveness of the approach has been evaluated. Our approach is computationally effective and easy to implement. Unlike population based methods, it takes into account the individual behavior of patients. Also, in contrast to individual based methods, it can put together consolidated information from the entire data to provide reliable initial estimates. In the next Section, the proposed method is extended to consider other types of disturbances.

2.3. Generalization of the Proposed Algorithm for No-Show and Cancellation Prediction

Algorithm 2 illustrates the pseudo code of the proposed algorithm for estimating two types of disruptions probabilities: no-show and cancellation. Similar to *Algorithm 1*, algorithm 2 consists of the following three main components:

1. Initial no-show and cancellation probabilities estimation
2. Bayesian update of the no-show and cancellation estimates
3. Weight optimization

Algorithm 2: No-show and Cancellation Prediction Algorithm

Input: Input data (X_{ij}, Y_{ij}), Threshold parameter T

Output: Estimated no-show and cancellation probabilities $\hat{p}^{Model}$, Dirichlet distribution posterior parameters (α_{ij}^{pos}), Multinomial logistic regression estimated parameters $\hat{B}_k$

Procedure:

```
/* Logistic regression*/
$\hat{B}_k \leftarrow$ Calculate MLE of Equation (1-5) parameter
$\hat{p}_{ik}^{0}(Y_i = 1,2,3|X_{ij}) \leftarrow F(X_{ij}, \hat{B}_k)$
$\alpha_{ik}^{pri} \leftarrow \hat{p}_{ik}^{0}$
/*Weight optimization*/
$\hat{p}_{i(j=v_2)k}^{Emp} \leftarrow \frac{\sum_{j=v_1}^{v_2} Y_{ijk}}{v_2 - v_1 + 1}$
Until Equation (2-1) improvement $D < T$ do
    $W$ ←set a value for the vector of appointments weights
    /*Bayesian update */
    $\alpha_{i(j=v_2)k}^{pos} \leftarrow \alpha_{ik}^{pri} + \sum_{i,j=v1}^{i,v2} \left( \prod_{\omega \in W} w_{ij\omega} \right) Y_{ij}$
    $\hat{p}_{i(j=v_2)k}^{Model} \leftarrow \frac{\alpha_{i(j=v_2)k}^{pos}}{\sum_{k=1}^{K} \alpha_{i(j=v_2)k}^{pos}}$
    $P_{Di(j=v_2)} \leftarrow \left[ p_{i(j=v_2)1}^{Emp} - \hat{p}_{i(j=v_2)1}^{Model}, \ldots, p_{i(j=v_2)K}^{Emp} - \hat{p}_{i(j=v_2)K}^{Model} \right]$
    $\bar{P}_D \leftarrow \frac{1}{n} \sum_{i=1}^{n} P_{Di}$
    $S_{P_D} \leftarrow \frac{1}{n-1} \sum_{i=1}^{n} (P_{D_i} - \bar{P}_D)'(P_{D_i} - \bar{P}_D)$
    $t^2 \leftarrow n \bar{P}_D S_{P_D}^{\ -1} \bar{P}_D$
    $F_0 \leftarrow \frac{n-K}{n(n-1)} t^2$
    $p - value_{paird\ F-test} \leftarrow p(F_0 > F_{a,p,n-K})$
Return $\hat{p}^{Model}$
```

In the first component, based on the training dataset consisting individuals' personal information (D_{GI}), (such as gender, marital status, etc.) and their sequence of appointment information (e.g. previous attendance records (D_{NR})), a multinomial logistic regression model $F\left(X_{ij}, \hat{B}\right)$ is formulated *(line 2)*. Then, using logistic regression, an initial estimate of no-show, cancellation and show-up probabilities are calculated, given by $\hat{p}_{ik}^{0}\left(Y_i = 1{,}2{,}3 | X_{ij}\right)$ *(line 3)*. As discussed in Section 1.1.2, Logistic regression bundles the information of the complete population together and finds a reliable initial estimate of no-show ($\hat{p}_{ik}^{0}$).

In the second component, which is interlaced with the third component, the initial estimate is used in a Bayesian update procedure to find the posterior no-show, cancellation and show-up probabilities of each person. For this purpose, $\hat{p}_{ik}^{0}$ is transformed into prior parameters of a Dirichlet distribution α_{ik}^{pri} as shown in *line 4*. Next, using the attendance record of each person $\left(Y_{ij}\right)$ the posterior parameters α_{ijk}^{Pos} and posterior probability of attendance $\hat{p}_{ijk}^{Model}$ is calculated (*lines 10 and 11*). As discussed in Section 1.1.3, the reason for applying the Bayesian update procedure to the output of logistic regression is that, typically regression models cannot consider individual patients behavior. Also, updating regression parameters based on new data records is both difficult and only marginally effective (especially when the model is already constructed on a huge dataset) in comparison to Bayesian update.

In the third component, appointments are weighted based on a subset of factors *W*= $[w_1, \ldots, w_\omega]$ (*line 8*) to increase the model performance in estimating the real probability of no-show. An optimization procedure is used for finding the optimal value of the weights. The objective function of the model is to minimize the difference between the empirical and estimated probabilities of no-show, cancellation and show-up as follows:

$$\begin{aligned} &\max p - value_{paired\ F-test} = p\left(F_0 > F_{\alpha,p,n-p}\right) \\ &\quad S.T: \\ &\quad 0 \leq w_q \leq 1 \qquad q = 1,..,\omega \end{aligned} \tag{2-4}$$

Where $w_1, \dots, w_\omega$ are the weights to be optimized and $p - value_{paird\ F-test}$ is the p-value of a one-sided statistical hypothesis testing of the paired estimated p using the model and estimated p using the attendance records:

$$\begin{cases} H_0 : p_D^{Model} = p_D^{\mathrm{Re}\,al} \\ H_1 : p_D^{Model} \neq p_D^{\mathrm{Re}\,al} \end{cases} \tag{2-5}$$

It should be noted that MSE can also be used as the objective function. However, F −statistic which is used above not only contains MSE information in itself (both S_P in the T^2 statistics and *MSE* estimate the deviation of a variable from its nominal value) (line 14), but also has a statistical distribution which makes it a better choice for our optimization model.

In Equation (2-4), $F_{a,K,n-K}$ is the percentage of points or value of F random variable with K and n-K degrees of freedom such that the probability that $F_{p,n-p}$ exceeds this value is α, and $F_0 = \frac{n-K}{n(n-1)} t^2$, where t^2 is calculated through *lines 11 -15*. $\hat{p}_{i(j=v2)k}^{Real}$ in line 11, shows the real rate of no-show, cancellation and showing up for person i calculated as $\hat{p}_{i(j=v2)k}^{Real} = \frac{\sum_{j=v1}^{v2} Y_{ijk}}{v2-v2+1}$ with Y_{ijk} as a multinomial (random) variable representing the records of attendance of type k for patient i and appointment j. Here, v_1 is the index of first appointment in the validation dataset which is discussed shortly, and v_1 is the index of last appointments in the validation dataset for patient i. Also, $\hat{p}_{i(j=v_2)k}^{Model} \leftarrow \frac{\alpha_{i(j=v_2)k}^{pos}}{\sum_{k=1}^{K} \alpha_{i(j=v_2)k}^{pos}}$ is the estimated probability of disruption of type k calculated based on weighted appointments in validation dataset using the proposed model.

The optimization procedure is as follows: at every iteration, a vector of weights is assigned to the appointments in validation dataset (*line 8*). The weighted appointments are then plugged into the Bayesian update mechanism for estimating the probability of no-show (*lines 10 and 11*). Next, the estimates of the proposed model and real attendance records are compared by forming a F-statistic (*lines 12 to 16*) and the p-value of the paired F-test (*line 17*) which shows the goodness of the assigned weights, is used for improving the initial set of weights (*line 7*) . This procedure continues until no improvement is observed. Then, the $\hat{p}^{Model}$ of the iteration resulted in the best value of objective function is used as the no-show estimate (*line 18*).

2.4. Experimental Results

For the purpose of evaluation, the performance of the proposed method is evaluated along with different population-based and individual-based algorithms on the extended version of the dataset used in the previous Section which includes 1,543 patient records with the following appointment information: (1) sex, (2) date of birth (DOB), (3) marital status, (4) medical service coverage, (4) address (zip code), (5) clinic and (6) prior attendance record in the hospital. 3-fold cross-validation with approximately 500 records each for training, validation and testing is considered for the evaluation.

We will first discuss the preprocessing of the data and then provide a stylized example for one sample patient record to illustrate how the model works. Finally, the result of applying the model on the dataset is discussed based on two types of analysis: (i) time-wise analysis and (ii) patient-wise analysis.

2.4.1. Data Preprocessing

As discussed in Section 2.2.1, before applying the proposed model to the dataset, it should be preprocessed (see Section 2.2.1 for more information). Table 6 shows the result of clustering the clinics based on their probability of no-show, cancellation and show-up using GMM. This Table is an extension to the Figure 1(b) in Section 2.2.1 . In the Section 2.2.1 the clinics are grouped based only one type of disruption, namely no-show rate. However, here they are clustered based on two types of disturbances, namely no-show and cancellation.

Table 6: The result of clustering the clinics based on their no-show, cancellation and show-up probabilities

		Cluster				
Attribute	**Parameter**	**0**	**1**	**2**	**3**	**4**
No-show	μ	0.0015	0.3488	0.2167	1.0000	0.1377
	σ	0.0114	0.1557	0.2692	0.3085	0.1331
Cancellation	μ	0.0011	0.0540	0.7833	0.0000	0.3276
	σ	0.0088	0.0733	0.2692	0.1968	0.1251
Show-up	μ	0.9974	0.5972	0.0000	0.0000	0.5346
	σ	0.0142	0.1523	0.0031	0.3372	0.1465

Again, like previous Section, (1) appointment recency, (2) appointment preceding non-work days (Saturday, Sunday, and holidays), and (3) clinic cluster, are considered as weighting factors ($W = [w_1, w_2, w_3]$), which are arranged in the data structure illustrated in Table 7. Table 7 also shows the optimal values of the weights used in Section 2.4.2 analysis.

Table 7: Data structure and optimal value of the weighting factors

Analysis	**Validation set**	**Appointment Recency**					**Preceding non-workday**		**Clinic Cluster**				
		<1 wk.	**<1 mon.**	**<3 mon.**	**6< mon.**	**> 6 mon.**	**Not-before holiday**	**Before holiday**	**Extremely important**	**Highly important**	**Moderately important**	**Low important**	**Not important**
time wise	1	1	1	1	0.75	0.5	1	0.90	1	1	0.9	0.81	0.62.
	2	1	0.97	0.96	0.82	0.67	1	0.98	1	1	0.84	0.73	0.71
	3	1	1	1	0.87	0.78	1	0.97	1	0.87	0.81	0.80	0.75
Patient wise	1	1	1	0.82	0.73	0.70	1	1	1	0.94	0.85	0.77	0.69
	2	1	0.99	0.75	0.61	0.58	1	0.89	1	1	0.81	0.69	0.66
	3	1	1	0.89	0.81	0.51	1	1	1	1	0.89	0.69	0.55

2.4.2. Applying the Proposed Model to a Sample Patient

The Section explains different steps of the proposed approach using a simple case study on a randomly selected patient. The patient was male, born on 02/5/1978, never married, degree of medical service coverage less than 5% with zip code 48235. Table 8 shows his appointment information as patterns of show/no-show from 10/13/2009 to 12/31/2009 (training data). Note that no-shows are represented by 1 while cancellation and show-ups are represented by 2 and 3 respectively.

Table 8: No-show, cancellation and show-up record of a sample patient

Appointment No.	Appointment date	Clinic cluster	Weight			Attendance record
			Preceding non-workday	Recency	Clinic cluster	
1	10/1/2009	0	1	0.5	0.9	3
2	10/8/2009	0	1	0.5	0.9	1
3	10/9/2009	0	0.9	0.5	0.9	2
4	10/13/2009	2	1	0.5	1	3
5	10/13/2009	2	1	0.5	1	3
6	10/13/2009	2	1	0.5	1	3
7	10/15/2009	0	1	0.5	0.9	3
8	10/15/2009	2	1	0.5	1	3
9	10/19/2009	2	1	0.5	1	1
10	10/19/2009	2	1	0.5	1	1
11	10/19/2009	0	1	0.5	0.9	1
12	10/22/2009	2	1	0.5	1	3
13	11/6/2009	2	0.9	0.5	1	3
14	11/6/2009	0	0.9	0.5	0.9	3
15	12/3/2009	2	1	0.5	1	3
16	12/18/2009	2	0.9	0.5	1	3

Using the patient's personal and appointment information as well as his previous attendance record, the parameters of the fitted multinomial logistic regression model are computed as shown in Table 9 (since we are modeling a categorical variable with three levels, namely no-show, cancellation and show-up, two sets of regression parameters are estimated (See Equation 1-5) .

Table 9: A sample multiple logistic regression model fitted to the dataset

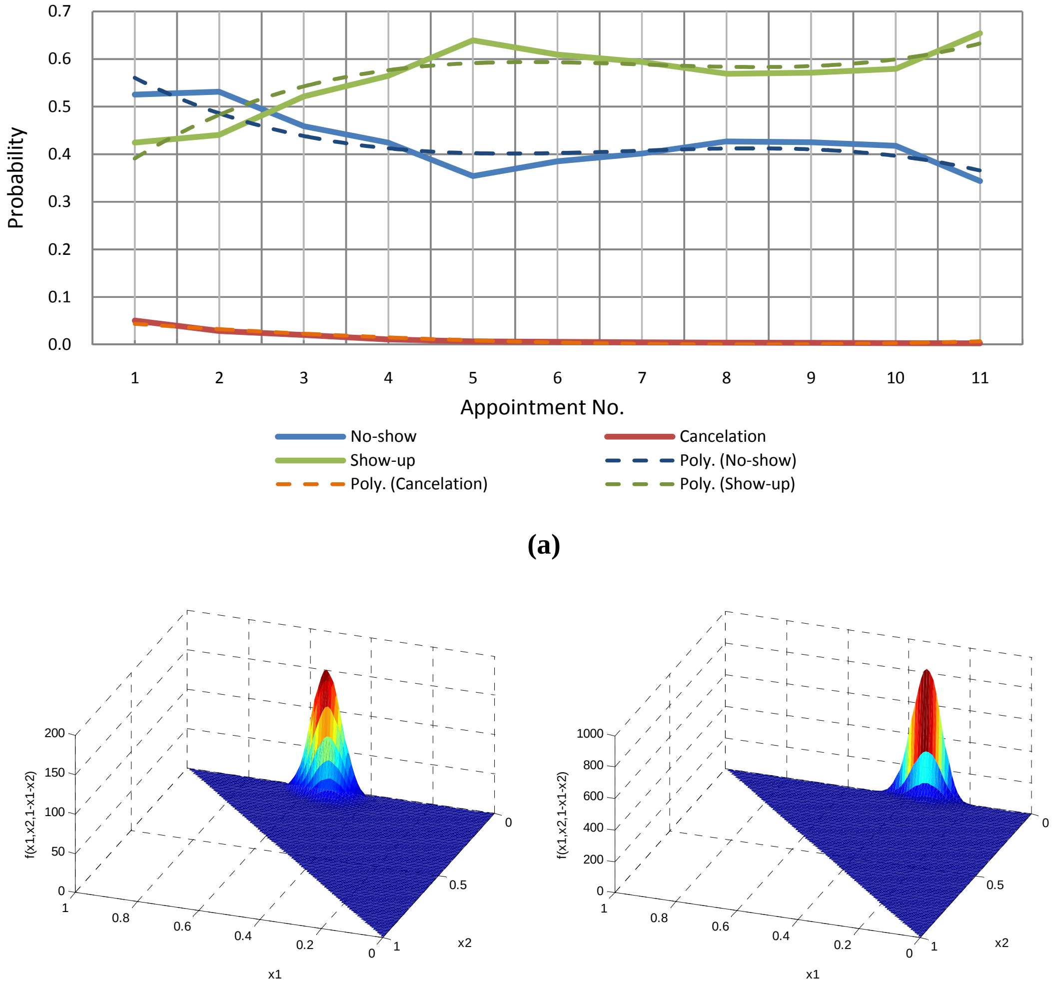
Sex	DOB	Marriage status	Medical service coverage	Zip code	Clinic cluster	Recency	Closeness to non-workday	Constant
106.605	-2.994	0.000	-0.314	-0.063	-0.002	-0.733	1.175	-0.388
168.919	-4.015	0.000	-0.234	-0.058	-0.003	-0.973	0.638	1.122

Based on the estimated coefficient the probability of no-show, cancellation and show-up for the first appointment in the testing dataset (2/1/2010) is estimated as (0,52525, 0.050374, 0.424376). This estimate is used for building the prior Dirichlet distribution with same parameters. Table 10 illustrates the updated parameters of Dirichlet distribution after each appointment.

Table 10: Bayesian update of Dirichlet distribution parameters

Appointment No.	Appointment date	Clinic category	Weight			Attendance record	Estimate Probability		
			Preceding non-work day	Recency	Clinic cluster		No-Show	Cancellation	Show-up
							0.5253	0.0504	0.4244
18	2/1/2010	2	1	0.75	1	1	0.7626	0.0252	0.2122
19	2/5/2010	2	0.9	0.75	1	3	0.3051	0.0101	0.6849
20	2/5/2010	2	0.9	0.75	1	1	0.4209	0.0084	0.5707
21	2/5/2010	0	0.9	0.75	1	3	0.2806	0.0056	0.7138
22	2/9/2010	2	1	0.75	1	1	0.3525	0.0050	0.6424
23	2/9/2010	2	1	0.75	1	1	0.4114	0.0046	0.5840
24	2/9/2010	2	1	0.75	1	1	0.4604	0.0042	0.5354
25	2/10/2010	2	1	0.75	1	1	0.5019	0.0039	0.4942
26	2/17/2010	2	1	0.75	1	3	0.4078	0.0031	0.5890
27	3/15/2010	2	1	0.75	1	3	0.3434	0.0027	0.6539

Figure 8 (a) illustrates the changes in the estimated probabilities of no-show, cancellation and show-up after each new record of attendance (solid lines) plus the estimated trend (using order three polynomials) of each type (dashed lines). Figure 8 (a) shows how the Bayesian update reacts quickly to each new data record, which means that the procedure can rapidly converge to the real distribution of no-show. Figure 8 (b) compares the prior and posterior distributions of this patient attendance probability before and after applying testing data. It is easy to follow the movement of the probability density function to the right and upper edges of the simplex (indicated by the arrows) which can be interpreted as decreasing the probability of cancellation significantly.



(a)

(b) **(c)**

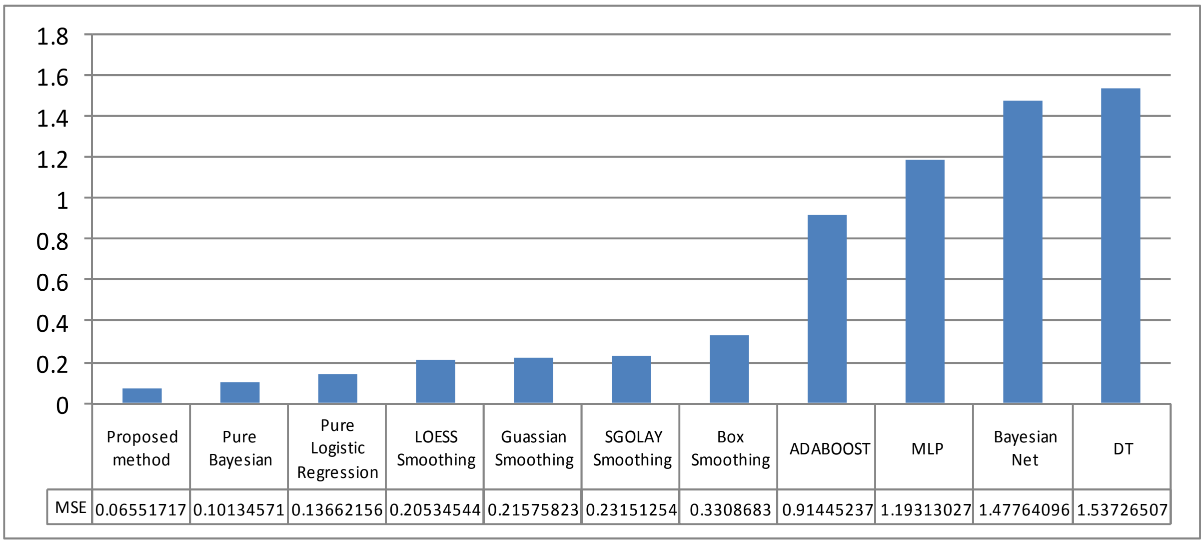
Figure 8: Applying the proposed model for a sample patient: (a) changing parameters during Bayesian update (b) prior distribution (c) posterior distribution

2.4.2.1. Time wise Analysis

This Section study the performance of the proposed model along with representatives from population-based and individual-based algorithms based on time-wise analysis. The training, validation and testing data are constructed as follows: appointments that occurred before 11/23/2009 have been used for training; appointments between 11/23/2009 and 2/1/2010 have been chosen for validation, and finally, appointments after 2/1/2010 have been considered for testing. The main reason for selecting the above dates is to have approximately equal number of data records each in the training, validation and testing datasets.

The methods used in our comparison are the following: locally weighted scatter plot (LOESS), Savitzky-Golay, Box, and Gaussian which are used as smoothing techniques (Simonoff [1996]), decision tree (DT), multiple logistic regression (with same predictors as used in the proposed model regression part) and pure multinomial Bayesian update, Bayesian Net, Multilayer Perceptron Neural Net (MLP) and a boosting algorithm. For setting the parameters of the comparison methods, the size of the moving window for Box smoothing was varied over the range of 1 to 7 and the optimal size (5) was considered for the comparisons. The standard deviation parameter of Gaussian distribution is experimented over 0.2 to 1; while 0.65 (the optimal value) was used. J48 was used for building the decision tree and ADABOOST PART [Viola and Jones 2002] algorithm is used for boosting method. For the pure Bayesian updating the Jeffery's prior $(0.33, 0.33, 0.33)$ is considered as the prior [Bolstad 2007]. For the multinomial logistic regression and smoothing methods the whole data set is used for building the model.

Figure 9 illustrates the MSE of the different methods used for comparisons. Based on the MSE measure, the proposed model outperforms other methods, while the rule-based method has the worst performance. As can be seen from the results, in general, individual-based methods outperform population-based methods, while bundling these methods together (as in our proposed method) significantly improves the overall performance.



	Proposed method	Pure Bayesian	Pure Logistic Regression	LOESS Smoothing	Guassian Smoothing	SGOLAY Smoothing	Box Smoothing	ADABOOST	MLP	Bayesian Net	DT
MSE	0.06551717	0.10134571	0.13662156	0.20534544	0.21575823	0.23151254	0.3308683	0.91445237	1.19313027	1.47764096	1.53726507

Figure 9: Mean Squared Error (MSE) of the studied methods for time-wise analysis

Figures 10-12 compare the empirical and estimated probability of no-show and cancellation for the methods over different patients (the performance of other methods along with the source code is available upon request). As shown in Figure 12, the proposed approach often predicts the real pattern correctly with considerable small variance.

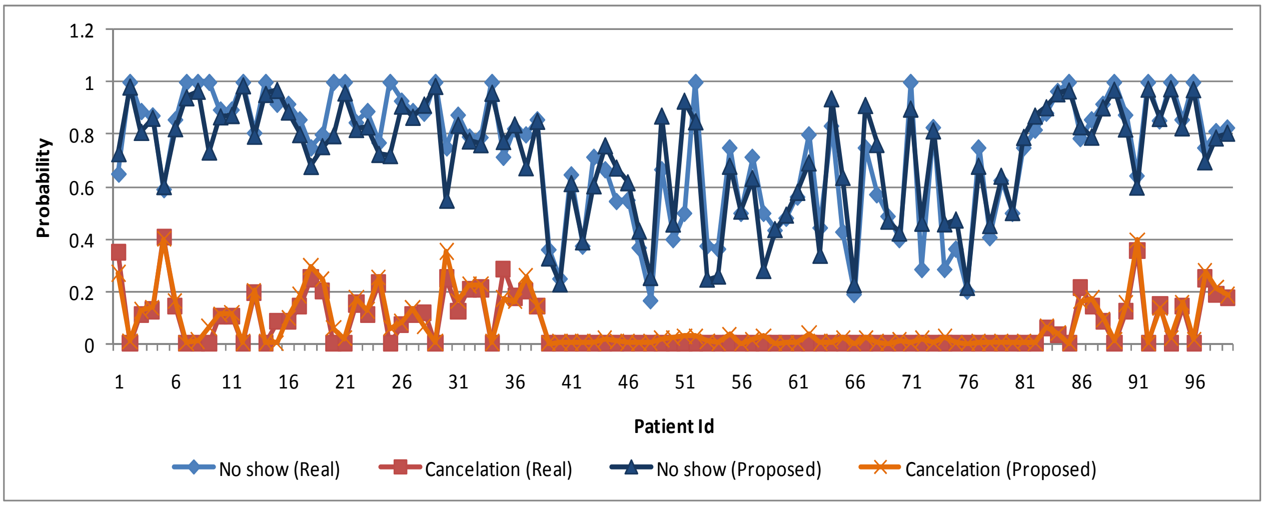


Figure 10: The performance of the proposed approach over different patients: estimated versus empirical probability of no-show and cancellation

Figure 11 illustrates the estimates from multinomial logistic regression, a population-based method. The estimates tend to have small fluctuations around an approximately fixed mean, though in general it somehow resembles the true pattern of the real no-show and cancellation. In addition, the difference between the estimated and real series significantly increases for patients with tendency of not cancelling their appointments (those could be either patients with good records of showing-up or those with high rate of no-show). Such result clearly shows that the regression models may not fully capture the difference among patients' personal behaviors.

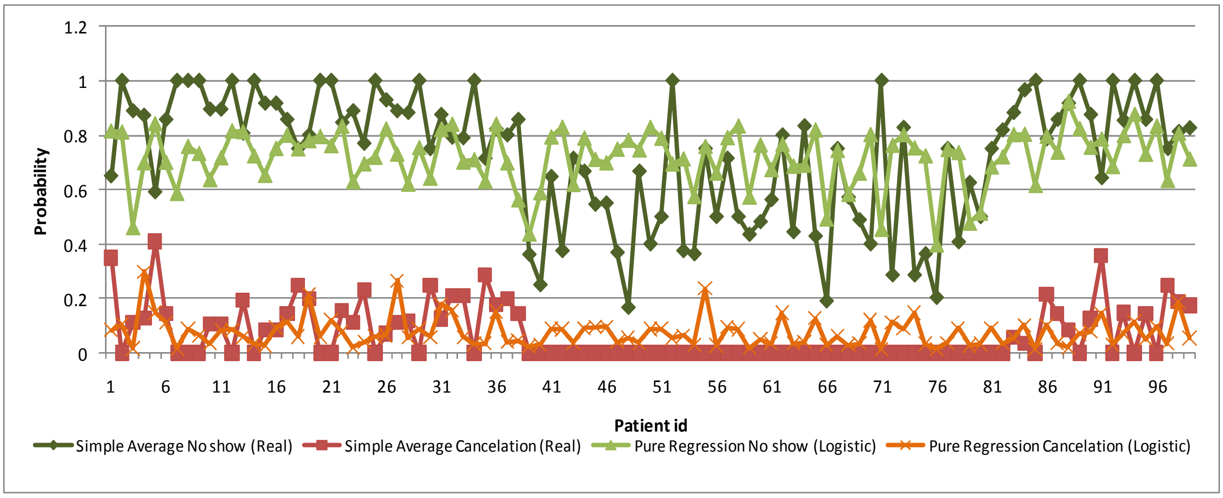


Figure 11: The performance of pure multinomial logistic regression over different patients: estimated versus empirical probability of no-show and cancellation

Finally, Figure 12 shows the results from pure Bayesian update method, which is a popular individual-based method. It can be seen that pure Bayesian update can basically detect the fluctuations in the real series correctly; however the estimates are far from the real ones in considerable number of cases. Further analysis revealed that such cases contain few number of attendance records which means that the pure Bayesian parameters update could not neutralize the effect of prior especially if that is far from the real case.

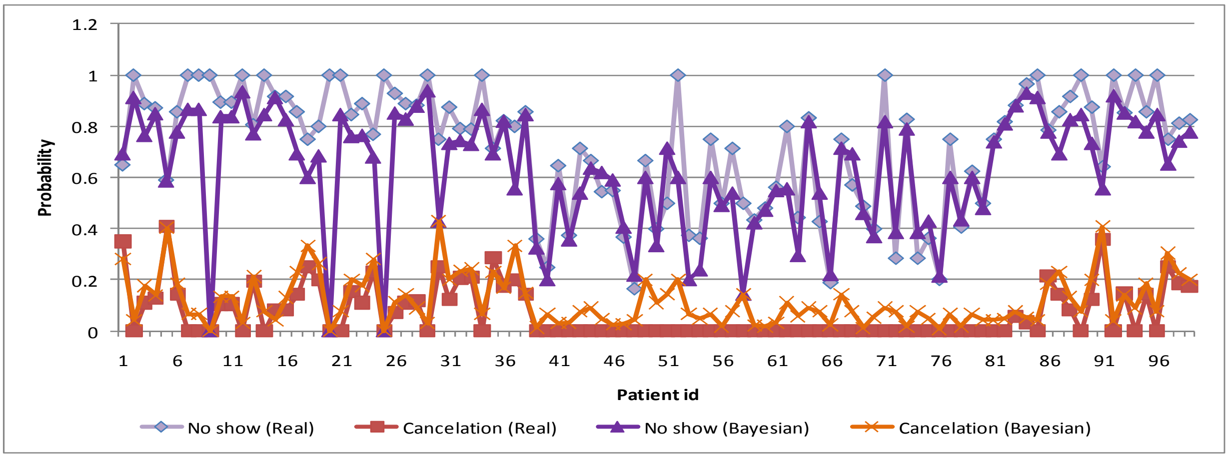


Figure 12: The performance of pure Bayesian updating method over different patients: estimated versus empirical probability of no-show and cancellation

2.4.2.2. Patient wise Analysis

Here the proposed method is compared to some other methods in the literature using patient-wise analysis. For this purpose, out of 99 patients in the database, using 3-fold cross-validation, 33 patients were randomly chosen for training, validation and testing. Figures 13-16 illustrate MSE and pairwise comparison of the empirical and estimated no-show and cancellation probabilities of the comparing methods which reveals similar results to the time-wise analysis discussed in the previous Section.

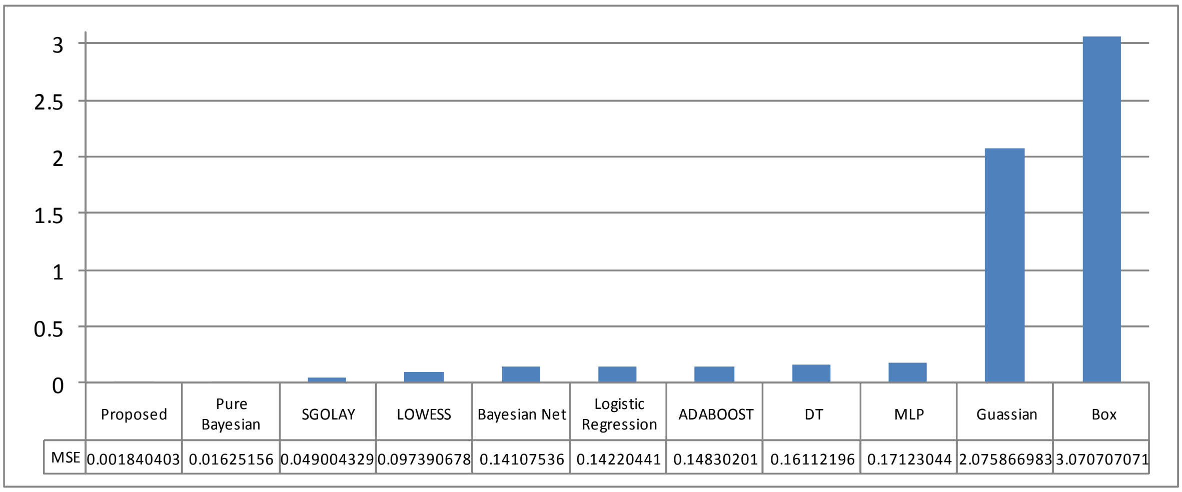


Figure 13: Mean Squared Error (MSE) of the studied methods for patient-wise analysis

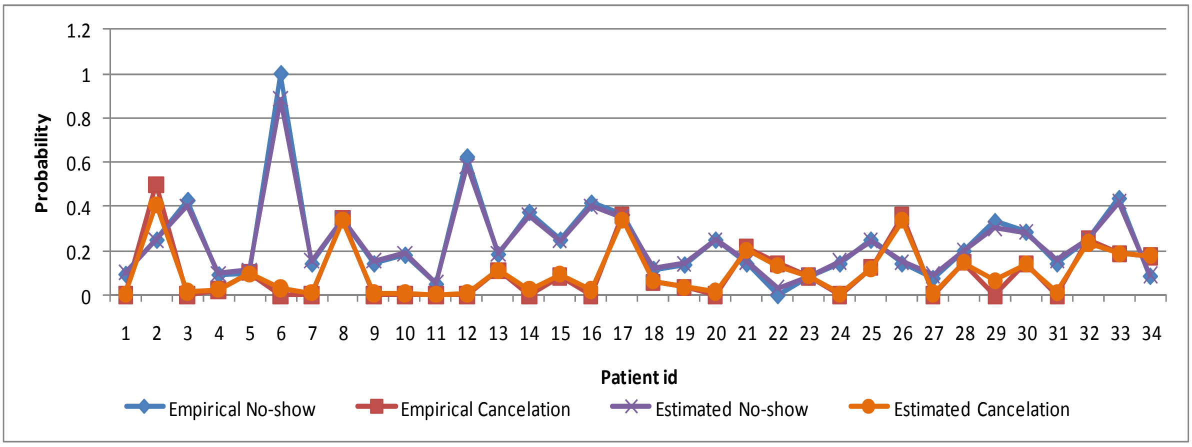


Figure 14: The performance of the proposed approach over different patients: estimated versus empirical probability of no-show and cancellation

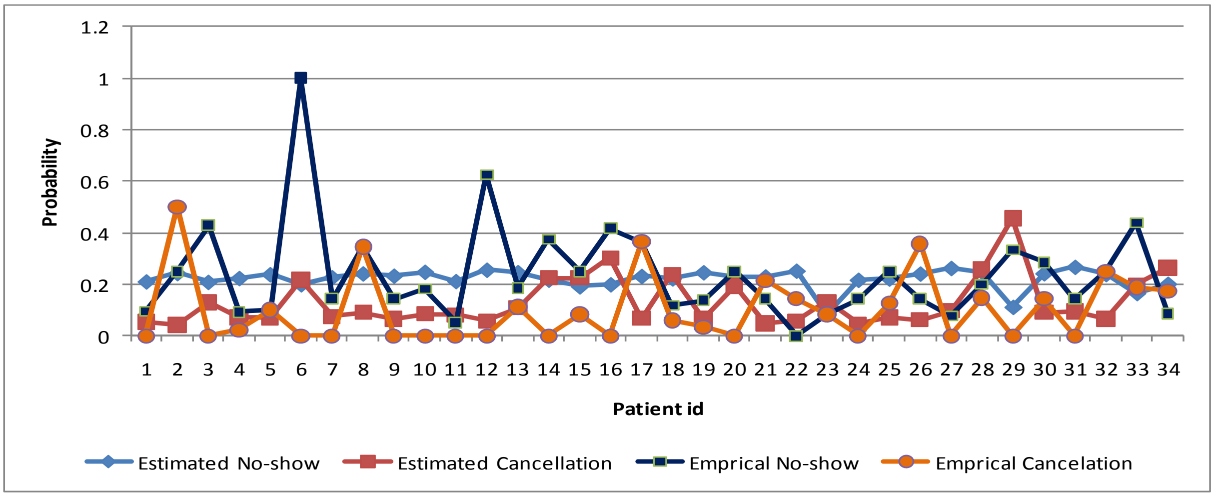


Figure 15: The performance of pure multinomial logistic regression over different patients: estimated versus empirical probability of no-show and cancellation

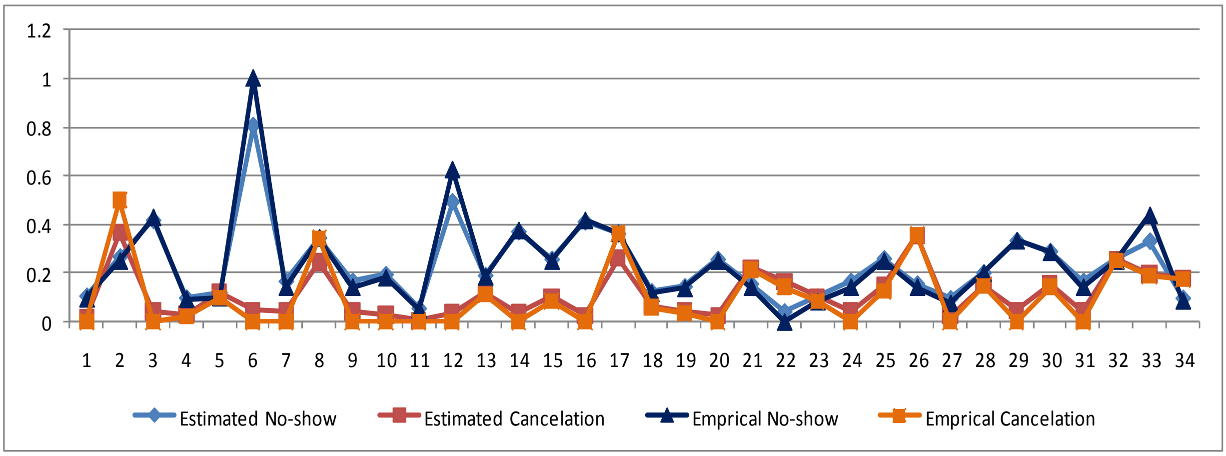


Figure 16: The performance of pure Bayesian updating method over different patients: estimated versus empirical probability of no-show and cancellation

The results from Figures 13-16 clearly show the capability of the proposed model in estimating probability of disruptions for both current and hypothetical patients of a health care system.

2.4.2.3. Discussion

In this Section, the probabilistic model in the previous Section has been extended to estimate the individuals' probabilities of no-show, cancellation and showing up in real-time. Also based on real-world patient data collected from a Veterans Affairs medical hospital, the effectiveness of the approach was evaluated. The result of the proposed method can be used to develop more effective appointment scheduling systems and more precise overbooking strategies to reduce the negative effect of no-shows and fill in appointment slots while maintaining short waiting times which is discussed in the next chapter.

CHAPTER 3 AN OPTIMIZATION MODEL FOR EFFECTIVE APPOINTMENT SCHEDULING IN THE PRESENCE OF DISTURBANCES

3.1. Introduction

Since the proposed prediction model has been developed for and implemented in a medical center, this Section discussed one of the most important applications of the proposed approach in health care systems which is scheduling. Though, the fields discussed in this Section can be extended to other service industries as well.

The estimate of the (attendance) disruptions probabilities can be effectively used in building robust scheduling systems [Cayirlit and Veral 2003; Gupta and Denton 2008]. A typical scenario in patient scheduling is that patients call a medical clinic to request an appointment with their physician. During the call, the scheduler assigns the patient to an available slot, a small time period (e.g. 30 minutes) in an operational or service period in the physician's schedule. This is communicated to the patient before the call terminates and hence, the schedule is constructed sequentially and has dynamic nature. Most of the existing models in the literature as well as the first part of this Section assume the demand is precisely known and consider the problem as a non-sequential optimization problem. However, few of the recent methods as well as the second part of this Section consider the dynamic nature of the problem as sequential optimization model [Muthuraman and Lawley 2008].

In practice, there is limited opportunity to adjust the schedule once the complete set of patients is known. However, scheduled patients might not attend or may cancel the appointment.The objective of the clinic is to schedule arriving appointment requests so that the long-run average expected net reward is maximized [Chakraborty et al. 2010]. The net reward is usually calculated as the expected revenue of serving patients at the scheduled time minus the expected cost of (1) system waiting including physician and staff waiting due to patients' no-show, cancellation, etc. and (2) system over time and patients over flow due to falling behind schedule, walk-in patients, etc.

Holding the common assumptions of independency in patients' attendance distribution, the patient scheduling problem can be modeled as a Markov Decision Process (MDP) [Liu et al 2009] which is an extension of Markov chains allowing choice and rewards [Feinberg and Shwartz 2002]. In such model the decisions are the choice of time slot assigned to each patient upon the appointment request, decision epochs are the times right after the appointment requests, and the system state at a decision epoch (which is a random variable affected by patients' behavior like no-show and cancellation) is the number of appointment requests as well as patients in the schedule at that decision epoch. Thus, accurate estimate of the individuals' probability of no-show and/or cancellation can result in a more an accurate schedule [Chakraborty et al. 2010; Hassin and Mendel 2008 ; and Liu et al. 2009]. Such accurate estimate can also be used for estimating the state of the scheduling system based on factors such as the probability of overflow, system waiting, etc.

The estimate of attendance disruption can also be used for developing selective overbooking strategies which is a vital component for improving patient access and stabilizing revenue when there is a significant chance that some scheduled patients will not show-up or cancel the appointment [Muthuraman and Lawley 2008]. Here, the main objective would be compensating for patient no-shows, and cancellation while maximizing the long-run average expected net reward. Because of no-shows, the clinic capacity will usually be underutilized without some overbooking. However, overbooking can reduce the negative effect of no-shows and fill in the appointment slots while maintaining short wait times [Muthuraman and Lawley 2008; and Zeng et al. 2010]. This problem can be modeled using multi-objective optimization with associated costs and rewards serving as weighting coefficients.

Finally, as another application of the proposed method which is not discussed in this study, the estimates of disruption probabilities can be used for designing proactive strategies such as reminder calls or some sort of penalty for encouraging high risk individuals, with elevated rate of no-show and cancellation [Ho and Lau 1992].

Table 11 describes the notation used in the proposed scheduling model.

Table 11: Notations used in the proposed scheduling model

Notations	**Description**
i,j	Indices of the time slot in a sample scheduling day ($i = 1, \dots I$), and the type of patients ($j = 1, \dots, J$)
λ	The parameter of exponential random variable for service time
$[D_{ij}]$	the estimate of demand for a specific date, such that D_{ij} represents the number of patients of type j (patients with no-show probability range *j*) requested an appointment for i^{th}slot of the day
$[S_{ij}]$	The matrix of schedule for a specific day, such that S_{ij} represents the number of patients of type j assigned for i^{th}slot of the day
$[R_{ik}]$	The overflow matrix for a specific day such that R_{ik} is the probability that *k* patients overflow from slot *i* to slot $i + 1$
$[Q_{ik}]$	The arrival matrix such that Q_{ik} is the probability that *l* patients arrive at the beginning of slot i
$[ORG_{ij}]$	The matrix of patients scheduled for slot *ij* of $[S_{ij}]$ who are scheduled in their desired time (slot *ij* of $[D_{ij}]$)
$[FROM_{ij}]$	The matrix of patients scheduled for slot *ij* of $[S_{ij}]$ while their desired slot in $[D_{ij}]$ is different
$[TO_{ij}]$	The matrix of patients in slot *ij* of $[D_{ij}]$ who assigned to a different slot in $[S_{ij}]$
$[ELIM_{ij}]$	The matrix of patient demands in $[D_{ij}]$ that are not considered in $[S_{ij}]$
r	The reward of serving a patient in its assigned slot in the schedule $[S_{ij}]$
c_i	The penalty of patient overflow from the scheduled slot i to the next slot (which can include the related overtime cost as well)
L_i	The number of patients receive service in slot i
T, t	The amount of time to the appointment date that the demands starts to be considered, and the time of receiving a demand (Hence, at each stage, $T - t$ is the remaining time to the appointment date)
$\gamma_{ij}(t)$	The rate of receiving demands for slot i of the schedule from a patient of type j

3.2. The Proposed Optimization Model for Non-Sequential Scheduling

Here a non-sequential optimization model will be developed for scheduling patients in a healthcare system while considering the effect of no-show (the proposed model can be extended to consider cancellation effect which is not being studied here). Also, for the economy of computational effort, the patients are categorized into J $(J \geq 1)$ $(j = \{1,2,\dots,J\})$ types by discrediting no-show probabilities. It is assumed that patients of type j_1 has less probability of no-show comparing to patient of type j_2 if $j_1 < j_2$. Furthermore, it is also assumed that each sample day in the schedule is divided into I $(I \geq 1)$ $(i = \{1,2,\dots,I\})$ time slots of equal length, each of which can take one or more patients depending on their service time which is assumed to be an exponential random variable with mean $\frac{1}{\lambda}$.

Having the estimate of demand $[D_{ij}]$ for a specific date, such that D_{ij} represents the number of patients of type j (patients with no-show probability range j) requested an appointment for i^{th}slot of the day, the net reward of a sample schedule S can be formulated as follows [Muthuraman and lawley 2008]:

$$\begin{aligned}
&\max G(S) = r\sum\nolimits_{i\in I} E[X_i] - \sum\nolimits_{i\in I} c_i E[Y_i] \\
&S.T: \\
&\sum\nolimits_{i\in I} S_{ij} \leq n \\
&S_{ij} \in Z \;\; \forall i \in I, j \in J
\end{aligned} \tag{3-1}$$

Where S_{ij} is the number of patient of type j that are assigned to slot i, X_i is a random variable representing the number of patients showing up at the beginning of slot i, and Y_i is another random variable denoting the number of patients who have not received service at the end of slot i . The relationship between X_i and Y_i can be shown as:

$$Y_i = \max\{Y_{i-1} + X_i - L_i, 0\} \tag{3-2}$$

Where L_i is the number of patients receive service in slot *i*. In Equation (3-2), *r* and c_i are the rewards of serving the patients in the assigned slot and the penalty of patient overflow to the next slot (which can include the related overtime cost as well) respectively. Therefore, the first term is the return from expected patient arrivals and the second term is the cost associated with the expected number of patients overflowed from one slot to another.

To compute probabilities for X_i and Y_i, Muthuraman and Lawley [2008] introduce two matrices, an arrival matrix $[Q_{il}]$ such that Q_{il} is the probability that l patients arrive at the beginning of slot i, and an overflow matrix $[R_{im}]$ such that R_{im} is the probability that *m* patients overflow from slot i to slot $i + 1$. These are computed as follows:

$$Q_{il}(S) = \Pr(X_i = l) = \sum\nolimits_{\pi\in\Omega}\prod\nolimits_{j\in J}\frac{S_{ij}!}{\pi_j!\left(S_{ij} - \pi_j\right)!} p_j^{\pi_j}\left(1 - p_j\right)^{S_{ij}-\pi_j} \tag{3-3}$$

where $\pi = \{\pi_1, \ldots, \pi_{|J|}\}$ with $\pi_j \in Z_+$ (the set of positive integers) for $j \in J, \sum_{j\in J}\pi_j = l$ and Ω is the set of all such vectors.

$$R_{im}(S) = \begin{cases} \sum_l \sum_k \left(l - F_{L_i}(l+k)\right) Q_{ik} R_{(i-1)l} & if\ m = 0 \\ \sum_l \sum_k f_{L_i}(l+k-m) Q_{ik} R_{(i-1)k} & if\ m \geq 0 \end{cases} \tag{3-4}$$

$$\begin{aligned} f_{L_i}(n) &= e^{-\lambda t}\frac{(\lambda t_i)^n}{n!} \\ F_{L_i}(n) &= \sum\nolimits_{\tilde{n}=0}^{n-1} f_{L_i}(\tilde{n}) \end{aligned} \tag{3-5}$$

where l is the number of patients arrive at the beginning of slot i and k is the number of patients overflow from slot $i - 1$ to slot$i$. Given these Equations one can compute $E[X_i] = \sum_l lQ_{il}$ and $E[Y_i] = \sum_k kR_{ik}$. Therefore, given $[D_{ij}]$ the objective of the model is to find the optimal schedule $\left[S_{ij}^{*}\right]$ (optimal assignment of patients to available time slots) which maximizes the net expected reward.

The model discussed in Equation (3-1) is capable of considering no-show probabilities and does overbooking which are the cases more frequently occur in practice. Yet, it has a few restricting assumptions some of which are given below:

1. Assigning patients to slots different than their desired time slot do not affect their no-show probability.
2. There is no walk-in patients (patients show-up without making an appointment)
3. The patient demand $\left[D_{ij}\right]$ is precisely known. Hence, the schedule can be made non-sequential.

In the rest of this Section, the first two of the above assumptions will be relaxed, and in the following Section the third assumption is considered.

3.3. An Optimization Model for Including changing no-show probabilities and walk-in Patients

In order to consider complex scenarios like walk-in patients and patient no-show probability changes with their assignment to undesired slot, as frequently occur in practice, it is necessary to know the decomposition from $\left[D_{ij}\right]$ to $\left[S_{ij}\right]$ which can be done by defining a set of four matrices with the same size of $\left[D_{ij}\right]$ and represented by $\left[ORG_{ij}\right], \left[FROM_{ij}\right], \left[TO_{ij}\right]$ and $\left[ELIM_{ij}\right]$. $\left[ORG_{ij}\right]$ shows patients scheduled at their desired time. $\left[FROM_{ij}\right]$ represents patients scheduled for slot ij of $\left[S_{ij}\right]$ while their desired slot was different. In contrast to $\left[FROM_{ij}\right]$, $\left[TO_{ij}\right]$ represents patients in slot ij of $\left[D_{ij}\right]$ who are assigned to a different slot in $\left[S_{ij}\right]$. Finally, $\left[ELIM_{ij}\right]$ shows patient demands in $\left[D_{ij}\right]$ not considered in $\left[S_{ij}\right]$. Using above matrices, $\left[S_{ij}\right]$

discussed in (21) can be gained by $[ORG_{ij}]+[FROM_{ij}]$, also $[D_{ij}]$ can be derived from $[ORG_{ij}]+[TO_{ij}]+[ELIM_{ij}]$. Therefore, the optimal schedule will be the one that optimizes all $[ORG_{ij}], [FROM_{ij}], [TO_{ij}]$ and $[ELIM_{ij}]$ matrices. To incorporate the above matrices into the optimization model, $Q_{il}(S)$ should be extended from binomial to multinomial as follows:

$$Q_{il}(S)=P_q(l)=P_q(X_{1i}+X_{2i}=l)=\binom{l}{X_{1i},X_{2i}}\Pr_1^{X_{1i}}\Pr_2^{X_{2i}} \tag{3-6}$$

where X_{1i} and X_{2i} are the total number of patients of different type $(j=1,..,J)$from $[ORG_{ij}]$ and $[FROM_{ij}]$ appears in slot i and $Pr_k(X_{ki})$ can be calculated using Equation (3-3). Based above extensions the optimization model can be rewritten as:

$$\begin{aligned}
&\max G(S)=r\left(\sum_{i\in I}E[X_{1i}]+\sum_{i\in I}E[X_{2i}]\right)-c_i\left(\sum_{i\in I}E[Y_{1i}]+\sum_{i\in I}E[Y_{2i}]\right)\\
&S.T:\\
&\sum_{i\in I}ORG_{ij}+\sum_{i\in I}FROM_{ij}\le n_j\\
&D_{ij}-(ORG_{ij}+TO_{ij}+ELIM_{ij})=0\\
&\sum_{i\in I}FROM_{ij}-\sum_{i\in I}TO_{ij}=0\\
&S_{ij}\in Z\ \forall i\in I, j\in J
\end{aligned} \tag{3-7}$$

The above model will consider changes in no-show probability by changing the patients' desired time. Furthermore, the proposed structure can handle other scenarios such as walk-in patients which would be resulted in following model:

$$
\begin{aligned}
&\max G(S)=r\left(\sum_{i\in I}E[X_{1i}]+\sum_{i\in I}E[X_{2i}]+\sum_{i\in I}\sum_{w\in W}E[X_{3i}]\right)-c_i\left(\sum_{i\in I}E[Y_{1i}]+\sum_{i\in I}E[Y_{2i}]\right)\\
&\quad S.T:\\
&\quad \sum_{i\in I}ORG_{ij}+\sum_{i\in I}FROM_{ij}\le n_j\\
&\quad D_{ij}-\left(ORG_{ij}+TO_{ij}+ELIM_{ij}\right)=0\\
&\quad \sum_{i\in I}FROM_{ij}-\sum_{i\in I}TO_{ij}=0\\
&\quad S_{ij}\in Z\ \forall i\in I, j\in J
\end{aligned}
\tag{3-8}
$$

As can be seen, considering walk-in patient can be done by simply adding another column to ORG_{ij}, $FROM_{ij}$,To_{ij} and $ELIM_{ij}$ matrices. The following example shows effectiveness of the proposed optimization model in scheduling patients in the presence of no-show.

3.3.1. Results on Simulated Data

To evaluate the performance of the proposed optimization model, this Section presents a simple scheduling problem for a single day of 8 hours with four 2 hour slots. Three types of patients has been considered as follows; (i) calling patients with show-up probability of 0.9, (ii) calling patients with show-up probability of 0.5 and (iii) walk-in patient with show-up probability of 0.2. $[D_{ij}]$ is considered to be known and equal to [4 2 1; 2 3 2; 3 3 2; 5 4 3] and the changing show-up probability of each patients of type 1, 2, and 3 when assigning them to slots different from their original request is considered as 0.7, 0.1 and 0.1 respectively. Also the reward of serving each patient at his assigned time is fixed at 200 and the cost of each patient overflow is set to 40.

Figure 17 compares the expected total profit of the proposed model and the original scheduling model without considering no-show (this model is simply obtained by setting all probabilities in the proposed model to one) under different no-show probabilities. As can be seen, the proposed approach uniformly performs better than the original model. In the meantime, when the uncertainty increases, the difference between the two methods gets more significant. The main reason of this difference is the ability of the proposed method for overbooking patients with high rate of no-shows in order to increase the system utility.

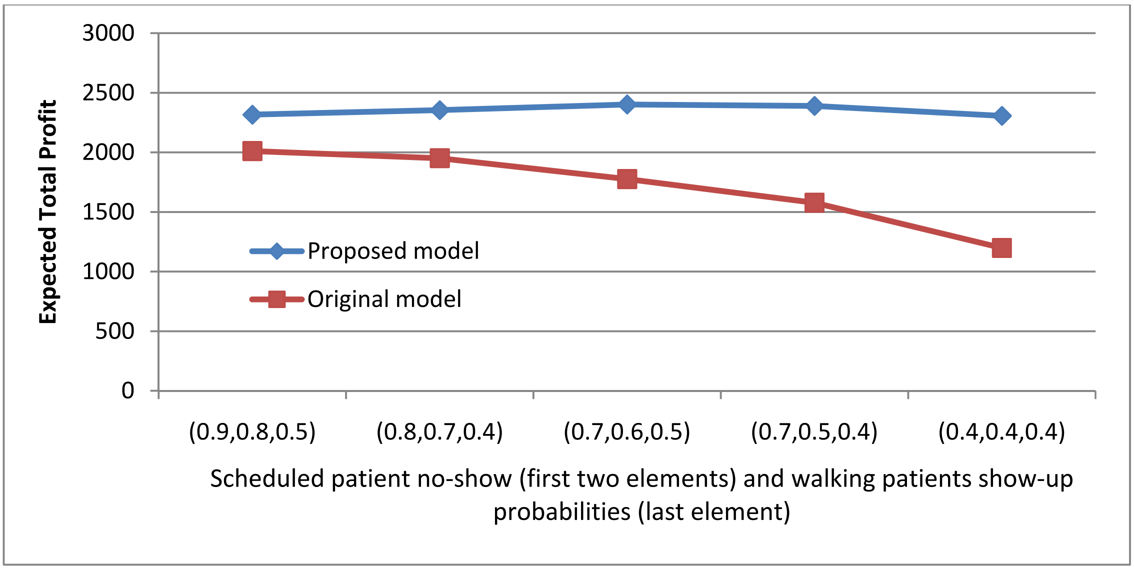


Figure 17: The expected total profit of the proposed non-sequential and the original scheduling model under different no-show probabilities for different types of patients

Figure 18 also illustrates the expected total profit of the proposed and the original model under different number of demands. As can be seen, again the proposed model outperforms the original model and the difference is increased as the number of demands grows.

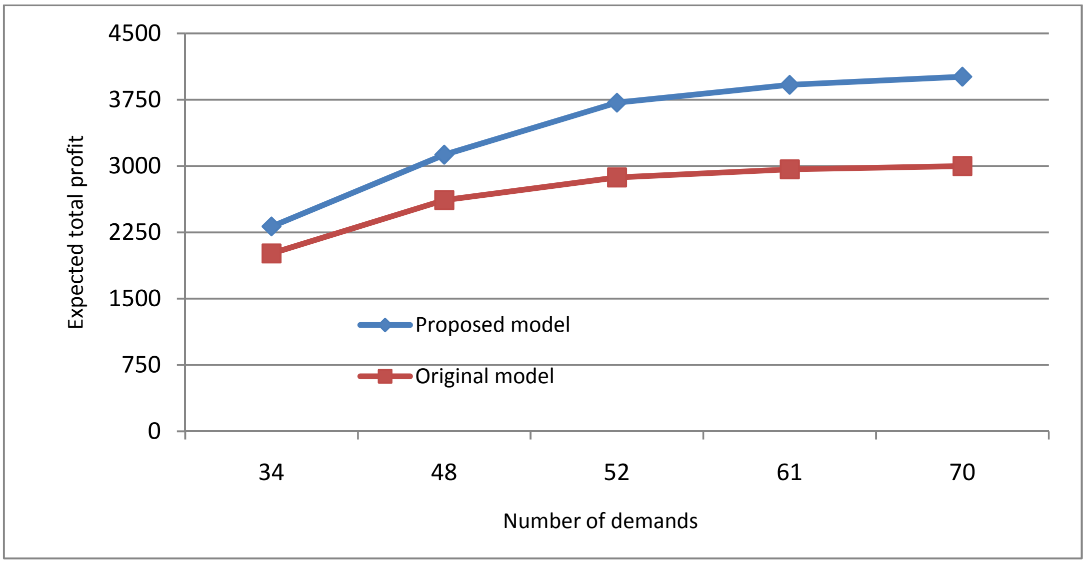


Figure 18: The expected total profit of the proposed non-sequential and the original scheduling model under number of demands

The result of Figures 17 and 18 shows that the proposed optimization model can effectively use the estimate of disturbances for overbooking patients with high rate of no-shows, which considerably increase the system performance under different situations.

3.4. Generalization of the Proposed Model for Sequential Scheduling

Here we will extend the model proposed in the previous Section to consider the case of sequential scheduling which occurs frequently in practice. The main difference between sequential and non sequential scheduling is:

1. Unlike non-sequential scheduling, in sequential scheduling during the call the scheduler assigns the patient to an available slot in the physician's schedule. This is communicated to the patient before the call terminates and, thus, the schedule is constructed sequentially.
2. Unlike non-sequential scheduling, in sequential scheduling the complete set of patients is not known when the schedule is generated. Schedulers do not know how many patients

will call for appointments and eventually be added to the schedule. Also, the schedulers do not know how many should be added, since they have no optimal stopping criteria.

3. Furthermore, there is little opportunity to adjust the schedule once completed.

Using the same matrix structure as described on the previous Section, the proposed optimization model for solving sequential scheduling problem can be written as follows:

$$
\begin{aligned}
&\max G(S)=r\left(\sum\nolimits_{i\in I}E[X_{1i}]+\sum\nolimits_{i\in I}E[X_{2i}]+\sum\nolimits_{i\in I}\sum\nolimits_{w\in W}E[X_{3i}]\right)-c_i\left(\sum\nolimits_{i\in I}E[Y_{1i}]+\sum\nolimits_{i\in I}E[Y_{2i}]\right)\\
&\quad S.T:\\
&\quad \sum\nolimits_{i\in I}ORG_{ij}+\sum\nolimits_{i\in I}FROM_{ij}\le n_j\\
&\quad D_{ij}-\left(ORG_{ij}+TO_{ij}+ELIM_{ij}\right)=0\\
&\quad \sum\nolimits_{i\in I}FROM_{ij}-\sum\nolimits_{i\in I}TO_{ij}=0\\
&\quad G\left(ORG_{ij}^{0}+FROM_{ij}^{0}+ORG_{ij}^{rem}+FROM_{ij}^{rem}+ORG_{ij}^{new}+FROM_{ij}^{new}\right)\\
&\qquad\quad >G\left(ORG_{ij}^{0}+FROM_{ij}^{0}+ORG_{ij}^{rem}+FROM_{ij}^{rem}\right)\\
&\quad S_{ij}\in Z\ \forall i\in I,j\in J
\end{aligned}
\tag{3-9}
$$

Where upper index 0 , e.g. ORG^0, represents elements of existing demand which already have been scheduled, new represents the just received demand which should be decided, and rem represents elements of estimated (remaining) demand related to future. Hence, $[ORG] = [ORG^0] + [ORG^{rem}] + [ORG^{new}]$, and $[FROM] = [FROM^0] + [FROM^{rem}] + [FROM^{new}]$.

As can be seen above, the new optimization model is gained by adding a constraint to the (3-8). This constraint guarantees that objective function should not decrease at any epoch (after each new decision on adding/rejecting a patient).

It should be noted that in the above model, we assume that ORG^0+$FROM^0$ shows the existing schedule and we have received a call from a patient of type j for slot i in a time that is T day ahead from the desired date of appointment. There is a simple strategy to check if the new patient should be added to the schedule and next to find the best slot that should be assigned to him/her.

First, the future demand for each slot of the desired date should be predicted for the remaining days. Next, the optimal assignment of the remaining demands should be calculated using Section Equation (3-8). Finally, it should be checked if the new request can be assigned to any available slot to which can improve the objective function (if there are multiple slots that can improve the objective function then the one with maximum improvement will be chosen). Finding a slot for the request is equivalent to accepting the request and vice versa. The following procedure shows the details of the proposed sequential optimization:

Step 1. Set $ORG_{ij}^{0} + FROM_{ij}^{0} = 0$ for all $i \in I$ and $j \in J$, $k = 1$ and $t = 0$.

Step 2. Wait for k^{th} patient call of type ij which is received at time t.

Step 3. Predict future patient demand for the remaining time to the appointment $T$and obtain $\widehat{D}_{ij}^{T-t}$.

Step 4. Using (3-8) find the optimal assignment of $\widehat{D}_{ij}^{T-t}$for the remaining demands $\left(ORG_{ij}^{rem} + FROM_{ij}^{rem}\right)$

Step 5. Find the objective function $G(ORG_{ij}^{0} + FROM_{ij}^{0} + ORG_{ij}^{rem} + FROM_{ij}^{rem})$ (this is the objective function for the case that we are not including the k^{th} patient)

Step 5. Check if adding the k^{th} patient call to any of the available slots $\left(ORG_{ij}^{new} + FROM_{ij}^{new}\right)$ can improve Step 5 objective function $\Big(G\big(ORG_{ij}^{0} + FROM_{ij}^{0} + ORG_{ij}^{rem} + FROM_{ij}^{rem} + ORG_{ij}^{new} + FROM_{ij}^{new}\big) > G\big(ORG_{ij}^{0} + FROM_{ij}^{0} + ORG_{ij}^{rem} + FROM_{ij}^{rem}\big)\Big)$. If so $ORG_{ij}^{0} = ORG_{ij}^{0} + ORG_{ij}^{new}$ and $FROM_{ij}^{0} = FROM_{ij}^{0} + FROM_{ij}^{new}$.

Step 6. If $G\left(ORG_{ij}^{0} + FROM_{ij}^{0} + ORG_{ij}^{rem} + FROM_{ij}^{rem} + ORG_{ij}^{new} + FROM_{ij}^{new}\right) >$ $G\left(ORG_{ij}^{0} + FROM_{ij}^{0} + ORG_{ij}^{rem} + FROM_{ij}^{rem}\right)$ for all $i \in I$ and $\in J$, stop. Otherwise, go to Step 2.

3.4.1. Estimation of Demand

One of the key elements of the proposed approach as well as the other scheduling system is the demand estimation. Unfortunately, most of the methods in the field of appointment scheduling are based on naïve methods such as averaging the previous demands. Here, a simple but innovative approach is proposed which is based on the survival analysis and Bayesian update for demand estimation.

Assuming that the demand for a specific day starts T days ahead (here we use days for the sake of simplicity, however other measure like hours, minutes, etc can also be used) the procedure can be summarized as follows:

1. Using the historical data and an appropriate survival model like Weibull distribution estimate $\gamma_{ij}(t)$ ($\gamma(t)$in here is equivalent to $\lambda(t)$ in survival analysis, the failure rate function)

2. Assuming there is $T - t$ days to the appointment date, after receiving a call from a patient of type j for slot i. Update the demand as follows:

$$D_{ij}^{Pos} = D_{ij}^{\Pr i} + 1 + \int_{t}^{T} t.\gamma_{ij}(t)dt \tag{3-10}$$

In above relation d_{ij}^{Pre}shows the total number of received calls before the current call for the interval $(0, t)$, 1 represents the current call, and $\int_t^T t.\gamma_{ij}(t)dt$ shows the expected demand for the remaining time.

3. Update $\gamma_{ij}(t)$ using the current call and Bayesian update mechanism

3.4.2. Results on Simulated Data

Here we extend our simple scheduling problem discussed in Section 3.3.1 for sequential scheduling. For this purpose, a period of 10 days each with 8 hours has been considered for receiving the calls for making the appointments for each single day like the example in the previous Section. The rate of incoming call is considered as $\lambda = 0.5$ per hour and the probability of having a high no-show rate, low no-show rate and walk-in patient is considered as [0.7 0.2 0.1]. The rest of the information remains the same as the ones in previous Section example.

Figure 19 compares the expected total profit of the proposed model and the myopic scheduling algorithm [Muthuraman and Lawley 2008], which is one of the few sequential scheduling algorithms in the literature, across a sequence of 40 patient calls.

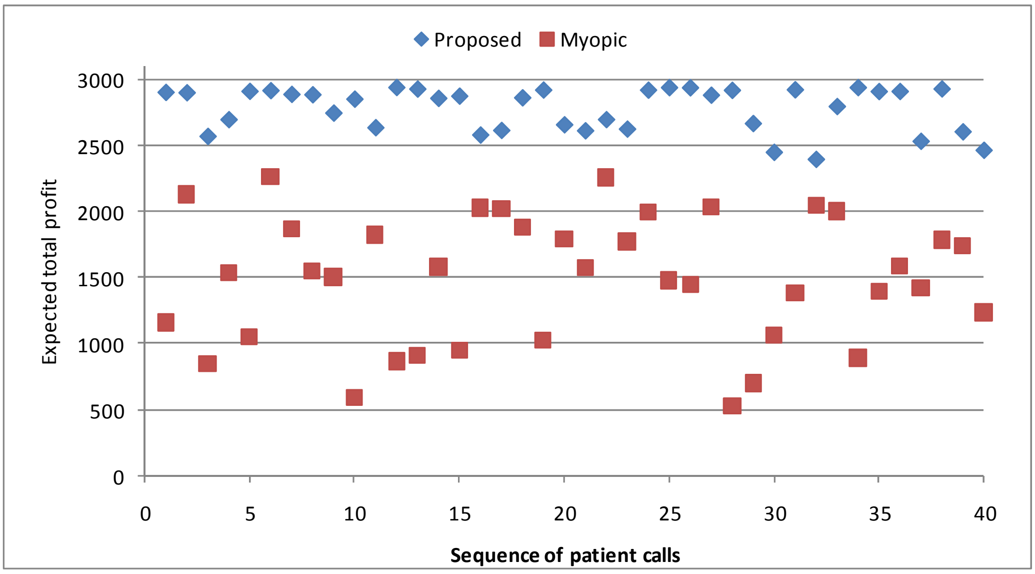


Figure 19: The expected total profit of the proposed sequential and the original scheduling model across a sequence of 40 patient calls

As can be seen the proposed approach always makes an upper bound of the myopic approach. It also has considerably smaller variance which clearly represents its effectiveness in modeling real-world situations.

CHAPTER 4 CONCLUSIONS AND FUTURE DIRECTIONS

Efficacy of any scheduling system depends highly on its ability to forecast and manage different types of disruptions and uncertainties. In this thesis, a probabilistic model based on logistic regression and Bayesian inference was developed to estimate the patients' disruption probability in real-time. A non-sequential and a sequential optimization model were also proposed which use disruption probabilities for appointment scheduling with overbooking strategy. Based on data collected at Veteran Affairs medical center, the effectiveness of the proposed prediction model was demonstrated. Furthermore, using two numerical synthetic examples the performance of the optimization models were evaluated in comparison with the common methods in the literature. The proposed prediction model is computationally effective and easy to implement. Unlike population based methods, our model takes into account the individual behavior of patients. Also in contrast to individual based methods it can accommodate vital information from the entire data collection and provide reliable initial estimates. In addition, the optimization model is flexible in formulating complex situations such as walk-in patients and changing no-show probabilities due to changing the patients' desirable time which occurs frequently in practice. The proposed prediction model can be easily extended to consider more sophisticated cases of disruptions and different types of prior distribution. Also the optimization model can be further extended to consider cancellations and delays.

Appendix A

Gaussian Mixture Models (GMM) and Expectation Maximization (EM) Algorithm

Gaussian Mixture Models (GMM) assume data points are drawn from a distribution that can be approximated by a mixture of Gaussian distributions. In this regard, assuming Q (the no-show rate of each clinic) is the feature vector, and k is the number of components (clinic clusters), the mixture model can be represented as follows [Reddy et al. 2008]:

$$p(Q \mid \Theta) = \sum_{i=1}^{k} a_i prob(Q \mid \theta_i) \tag{A-1}$$

Where $\Theta = \{a_1,...,a_k,\theta_1,...,\theta_k\}$ is the collection of parameters with $0 \le a_i \le 1, \forall i = 1,2,...,k$ and $\sum_{i=1}^{k} a_i = 1$ and $p(Q \mid \theta_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{Q-\mu_i}{2\sigma_i^2}\right)$. Having as a set of *n*, i.i.d samples $Q = \{q^{(1)}, q^{(2)},..., q^{(n)}\}$ from the above model the log-likelihood function can be rewritten as:

$$\log p(Q \mid \theta_i) = \log \prod_{j=1}^{n} p(q^{(j)} \mid \Theta) = \sum_{j=1}^{n} \log \sum_{i=1}^{k} \alpha_i p(q^{(j)} \mid \theta_j) \tag{A-2}$$

Here, the goal is to find Θ that maximizes the log-likelihood function:

$$\hat{\Theta}_{MLE} = \arg \max_{\Theta} \{\log p(Q \mid \Theta)\} \tag{A-3}$$

The surface of the above likelihood function is highly nonlinear, and no closed form solution exists for the above likelihood function. One way to deal with this problem is by introducing a hidden variable Z:

$$\log p(Q, Z \mid \theta_i) = \sum_{j=1}^{n} \sum_{i=1}^{k} z_i^{(j)} \log\left[\alpha_i p(q^{(j)} \mid z_i^{(j)} \theta_j)\right] \tag{A-4}$$

and using Expectation Maximization (EM) algorithm as follows [33]:

i. Initializing parameters Θ

ii. Iterating the following until convergence:

E-Step:

$$Q\left(\Theta \mid \Theta^{(t)}\right) = E_z \log\left[p\left(Q, Z \mid \Theta\right) \mid \Theta^{(t)}\right] \tag{A-5}$$

M- Step:

$$\Theta^{(t+1)} = \mathrm{argmax} Q\left(\Theta \mid \Theta^{(t)}\right) \tag{A-6}$$

REFERENCES

1. Atchison, J. 1985. Practical Bayesian problems in simplex sample spaces. *Bayesian Statistics* 2, 15-31. North-Holland/Elsevier (Amsterdam; New York).
2. Agresti, A. 2002. *Categorical Data Analysis*. Wiley-Interscience., New York, USA.
3. Alpaydin, E. 2010. *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*, Second Edition, MIT Press, Massachusetts, USA.
5. Bean, A.G., and Talaga, J., 1995. Predicting appointment breaking. *J Health Care Mark,* 15, 1, 29–34.
6. Bech, M. 2005. The economics of non-attendance and the expected effect of charging a fine on non-attendees. *Health Policy*. 74, 2, 181-191.
7. Barron, W.M. 1980. Failed Appointments: who misses them, why they are missed, and what can be done. *Primary Care*, 7, 4, 563–574.
8. Bolstad, W.M. 2007. *Introduction to Bayesian Statistics*. Wiley-Interscience, New York, USA.
9. Brockwell, P.J., and Davis R.A. 2009. *Time Series: Theory and Methods (Springer Series in Statistic)*. Springer, New York, USA.
10. Campbell, J.D., Chez, R.A., Queen, T.B.A., and Patron, E.,2000. The no-show rate in a high-risk obstetric clinic, *Journal of Women's Health & Gender-Based Medicine,* 9, 8, 891-895.
11. Cashman, S.B., Savageau, J.A., Savageau, L., Celeste , A., and Ferguson, W. 2004. Patient health status and appointment keeping in an urban community health center, *Journal of Health Care for the Poor and Underserved,* 15, 474–488.

12. Cayirli, T., and Veral, E. 2003. Outpatient scheduling in health care: A review of the literature. *Prod Opns Mngt*. 12, 4, 519–549.
13. Chakraborty, S., Muthuraman, K., and Lawlwey, M. 2010. Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*. 42, 5, 354-366.
14. Cote, M.J., 1999. Patient flow and resource utilization in an outpatient clinic. *Socio-Econ Plan Sci*. 33, 231–245.
15. Cynthia T. R., Nancy H. G., Clark S., Donna S.J., Wallace, D.W., and Arch P.A. 1995. Patient appointment failures in pediatric resident continuity clinics, *Pediatric and Adolescent medicine*, 149, 6, 693-695.
16. Dove H.G., and Karen C.S.. 1981. The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics, *Medical Care*, 19, 7, 734-740
17. Dove, H.G., and Schneider, K.C., 1981. The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics. *Med Care*. 19, 7, 734–740.
18. Dreihera, J, Froimovicia, M., Bibia, Y., Vardya, D.A., Cicurela, A., and Cohen, A.D. 2008, Nonattendance in obstetrics and gynecology patients, *Gynecol Obstet Invest*, 66, 40-43.
19. Evans, M., Hastings, N., and Peacock, B. 2000. *Statistical Distributions*, Third Edition, Wiley-Interscience, New York, USA.
20. Feinberg, E.A., Shwartz, A. 2002. *Handbook of Markov decision processes: methods and applications*. Kluwer Academic Publishers, Massachusetts, USA.
21. Forster, J.J., Skene, A.M. 1994. Calculation of marginal densities for parameters of multinomial distributions. *Statistics and Computing*, 4, 279-286.

22. Garuda, S.R., Javalgi, R.G., and Talluri , V.S. 1998, Tackling no-show behavior: A market driven approach. *Health Mark Q*, 15, 4, 25–44.
23. Glowacka K.J., Henry R.M., and May J.H. 2009. A hybrid data mining/simulation approach for modeling outpatient no-shows in clinic scheduling. *Journal of the Operational Research Society*, 60, 1056 -1068.
24. Goldman, L., Freidin, R., Cook, E.F., Eigner, J., and Grich, P. 1982. A multivariate approach to the prediction of no-show behavior in a primary care center, *Arch Intern Med,* 142, 563–567.
25. Goutis, C. 1993. Bayesian estimation methods for contingency tables, *Journal of the Italian Statistical Society* 2. 35-54.
26. Gupta, D., and Denton, B. 2008. Appointment scheduling in health care: challenges and opportunities. *IIE Transactions*, 40, 800–819.
27. Hassin, R., and Mendel, S. 2008. Scheduling arrivals to queues: A single-server model with no-shows. *Management Science*. 54, 3 , 565–572.
28. Haupt, R.L. and Haupt, S.E. 2004. *Practical Genetic Algorithms*. Second Edition, John Wiley, New York, USA.
29. Hixon, A.L., Chapman, R.W. and Nuovo, J. 1999. Failure to keep clinic appointments: implications for residency education and productivity. *Family Medicine* 31 (9), 627-630.
30. Ho, C. and Lau, H. 1992. Minimizing total cost in scheduling outpatient appointment. *Management Science*, 38 , 2 , 1750–1764.
31. Laganga L.R., and Lawrence, S.R. 2007. Clinic overbooking to improve patient access and increase provider productivity. *Decision Science*, 38, 251–276.

32. Lehmann, T.N.O., Aebia, A., Lehmann, D., Balandraux, O.M., and Stalder H. 2007. Missed appointments at a Swiss university outpatient clinic. *Public Health*. 121, 10, 790-799.
33. Leonard, T., 1973. A Bayesian method for histograms, *Biometrika* 60, 297-308.
34. Liu, N, Ziya, S, and Kulkarni, VG. 2009. Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management*, 12, 347-364.
35. Muthuraman, K., and Lawley, M. 2008. A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions*, 40, 9, 820 – 837.
36. Moore, C.G.,Wilson-Witherspoon, P., and Probst, JC. 2001. Time and money: Effects of no-shows at a family practice residency clinic. Family Medicine, 33, 7, 522–7.
37. Reddy C. K., Chiang, H.D. and Rajaratnam, B. 2008. TRUST-TECH-based expectation maximization for learning finite mixture models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 7, 1-11.
38. Rust C. T., Gallups N. H., Clark S., Jones D.S., Wilcox W.D., and Adolesc. A. 1995. Patient appointment failures in pediatric resident continuity clinics. *Pediatric and Adolescent medicine*, 149, 6, 693-695.
39. Simonoff J.S. 1996. *Smoothing Methods in Statistics (Springer Series in Statistics)*, Springer, New York, USA.
40. Viola P., Jones. M. 2002. Fast and robust classification using asymmetric AdaBoost and a detector cascade. *In Proceedings of Neural Information Processing Systems (NIPS)*.
41. Zeng B.O., Turkcan A., and Lin J. 2010. Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research*. 178, 1, 121-144.

ABSTRACT

PROBABILISTIC MODELS FOR PATIENTS SCHEDULING

by

ADEL ALAEDDINI

May 2011

Advisor: Dr. Chandan K. Reddy

Major: Computer Science

Degree: Master of Science

In spite of the success of theoretical appointment scheduling methods, there have been significant failures in practice primarily due to the rapid increase in the number of no-shows and cancelations from the individuals in recent times. These disruptions not only cause inconvenience to the management but also has a significant impact on the revenue, cost and resource utilization. In this research, we develop a hybrid probabilistic model based on logistic regression and Bayesian inference to predict the probability of no-shows in real-time. We also develop two novel non-sequential and sequential optimization models which can effectively use no-show probabilities for scheduling patients. Our integrated prediction and optimization model can be used to enable a precise overbooking strategy to reduce the negative effect of no-shows and fill appointment slots while maintaining short wait times. Using both simulated and real-world data, we demonstrate the effectiveness of the proposed hybrid predictive model and scheduling strategy compared to some of the well-studied approaches available in the literature.

AUTOBIOGRAPHICAL STATEMENT

Adel Alaeddini is a Ph.D. candidate of Industrial Engineering and M.S. student of computer science at Wayne State University. He used to work as a lecturer in the Departments of Industrial Engineering, and Information Technology of Azad University, Tehran, Iran. His main areas of research are statistical learning, global optimization, response surface methodology and quality engineering. He has two nominations for the best paper award in top conferences, three co-authored grants, more than ten published journal papers and several refereed conference papers. Based on his Master's thesis work, he published the following two articles:

i. Alaeddini, A., Reddy, Ch.K., and Yang, K. 2011. Predicting disturbances in appointment scheduling through hybrid probabilistic modeling. *ACM Transactions on Intelligent Systems and Technology*. Under Review.

ii. Alaeddini, A., Yang, K., YU, S., and Reddy, C.K. 2011. A probabilistic model for predicting the rate of no-show in hospital appointments, Healthcare management Science, (2011), To Appear.